Statistical Inference under Adaptive Sampling with LinUCB

Wei Fan*, Kevin Tan*, Yuting Wei

Department of Statistics and Data Science
The Wharton School, University of Pennsylvania

November 30, 2025

Abstract

Adaptively collected data has become ubiquitous within modern practice. However, even seemingly benign adaptive sampling schemes can introduce severe biases, rendering traditional statistical inference tools inapplicable. This can be mitigated by a property called stability, which states that if the rate at which an algorithm takes actions converges to a deterministic limit, one can expect that certain parameters are asymptotically normal. Building on a recent line of work for the multi-armed bandit setting, we show that the linear upper confidence bound (LinUCB) algorithm for linear bandits satisfies this property. In doing so, we painstakingly characterize the behavior of the eigenvalues and eigenvectors of the random design feature covariance matrix in the setting where the action set is the unit ball, showing that it decomposes into a rank-one direction that locks onto the true parameter and an almost-isotropic bulk that grows at a predictable $\sqrt{T}$ rate. This allows us to establish a central limit theorem for the LinUCB algorithm, establishing asymptotic normality for the limiting distribution of the estimation error where the convergence occurs at a $T^{-1/4}$ rate. The resulting Wald-type confidence sets and hypothesis tests do not depend on the feature covariance matrix and are asymptotically tighter than existing nonasymptotic confidence sets. Numerical simulations corroborate our findings.

Contents

---

*Equal contribution.

1 Introduction

Statistical inference for adaptively collected data is essential for providing rigorous justification and interpretability in modern data analysis, with applications ranging from scientific discovery to social decision-making. In sharp contrast to classical i.i.d. settings, adaptive data collection induces intricate dependencies across samples, often rendering traditional inferential tools unreliable. It is now well recognized that even seemingly benign adaptive sampling schemes can introduce severe biases and complicate the asymptotic distribution of estimators (see, e.g. Dickey and Fuller (1979); Lai and Wei (1982); Deshpande et al. (2023)), which in turn complicates the task of uncertainty quantification (Deshpande et al. (2018); Khamaru and Zhang (2024); Lin et al. (2023); Zhang et al. (2020, 2021)).

A central theme emerging from this growing body of work is that the very process of learning dynamically reshapes the statistical properties of the data. Although this complicates the analysis, recent work within the multi-armed bandit setting (Kalvit and Zeevi, 2021; Khamaru and Zhang, 2024; Han et al., 2024; Halder et al., 2025) has shown that certain algorithms exhibit a notion of "stability" that allows for asymptotic normality of the arm mean reward estimates. In other words, if the rate at which a multi-armed bandit algorithm pulls each arm is asymptotically deterministic, then under suitable conditions this alone can ensure that the estimated mean rewards are asymptotically normal. This property is satisfied for the UCB algorithm (Khamaru and Zhang, 2024; Kalvit and Zeevi, 2021; Han et al., 2024), but not Thompson sampling (Zhang et al., 2021) unless the posterior variance is inflated by a logarithmic factor (Halder et al., 2025).

We explore whether this phenomena of stability for the UCB algorithm (Khamaru and Zhang, 2024; Kalvit and Zeevi, 2021; Han et al., 2024) also extends to the linear bandit problem, a classical and influential model in reinforcement learning and the bandit literature. The linear bandit formalizes sequential decision-making in which the expected reward is a linear function of an action's features, making it both a natural and useful abstraction. Mathematically, given a content $\boldsymbol{x}_t \in \mathcal{X}$, and an action $\boldsymbol{a}_t \in \mathcal{A}$ at time t, the learner receives the reward

$$r_t = \langle \boldsymbol{\phi}(\boldsymbol{x}_t, \boldsymbol{a}_t), \boldsymbol{\theta}^\star \rangle + \epsilon_t \in \mathbb{R}, \tag{1}$$

according to an unknown parameter $\boldsymbol{\theta}^\star \in \mathbb{R}^d$. Here, $\boldsymbol{\phi} : \mathcal{X} \times \mathcal{A} \to \mathbb{R}^d$ is a feature map, and the action $\boldsymbol{a}_t \in \mathcal{F}_{t-1}$ with $\mathcal{F}_{t-1}$ being the σ-field generated by the history including previous actions and rewards $\{\boldsymbol{a}_1, r_1, \ldots, \boldsymbol{a}_{t-1}, r_{t-1}\}$, collected up to time $t-1$. The noise satisfies $\mathbb{E}[\epsilon_t | \mathcal{F}_{t-1}] = 0$.

Within this linear bandit framework, we focus on the linear upper confidence bound algorithm (*LinUCB*) of Li et al. (2010); Abbasi-yadkori et al. (2011), a canonical UCB-type method tailored to linear bandits. LinUCB stands out for its principled balance between exploration and exploitation and for its strong practical performance. Yet while LinUCB's regret guarantees are well understood, the distributional behavior of its estimators—and, consequently, tools for valid statistical inference under LinUCB—remain underdeveloped. A common fallback is the familiar non-asymptotic confidence set for $\boldsymbol{\theta}^\star$ (also used for action selection), but because it depends on the empirical feature covariance $\boldsymbol{\Lambda}_T$, it provides neither a limiting distribution for the estimator nor a deterministic characterization of the set's width. Beyond this stopgap, only a handful of results from stochastic approximation or stochastic gradient descent (Polyak and Juditsky, 1992; Su and Zhu, 2023; Wu et al., 2025; Chen et al., 2020) and a covariance characterization within LinUCB (Banerjee et al., 2023) speak to inference, and these remain insufficient for conducting valid statistical inference with LinUCB. This gap motivates the following question:

Can LinUCB be used not only as a learning algorithm, but also as a vehicle for valid statistical inference?

1.1 Prior art

Existing results do not provide a complete answer to this question. It has long been known that the confidence sets constructed by LinUCB contain $\boldsymbol{\theta}^*$ with high probability at each timestep. More precisely, when $\|\boldsymbol{\phi}(\boldsymbol{x}_t, \boldsymbol{a}_t)\|_2 \leq L$ for all $t = 1, ..., T$, $\|\boldsymbol{\theta}^\star\|_2 \leq S$ and $\boldsymbol{\Lambda}_t$ is the feature covariance matrix at time t (defined in (7)), Theorem 2 of Abbasi-yadkori et al. (2011) states that, given some regularization parameter $\lambda > 0$, for all $t = 1, ..., T$, with probability at least $1 - \delta$:

$$\boldsymbol{\theta}^* \in \mathcal{C}_t, \qquad \mathcal{C}_t = \left\{ \boldsymbol{\theta} \in \mathbb{R}^d \ \middle| \ \|\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}\|_{\boldsymbol{\Lambda}_t} \leq \sigma\sqrt{d\log\left(\frac{1 + TL^2/\lambda}{\delta}\right)} + \lambda^{1/2}S \right\}, \tag{2}$$

where $\|\boldsymbol{a}\|_{\boldsymbol{\Lambda}_t} := \sqrt{\boldsymbol{a}^\top \boldsymbol{\Lambda}_t \boldsymbol{a}}$. This yields a simultaneous confidence set for all coefficients within the parameter $\boldsymbol{\theta}^*$ with non-asymptotic $1-\delta$ coverage. However, this simply states that $\boldsymbol{\theta}^*$ is contained within some ellipsoid centered at $\widehat{\boldsymbol{\theta}}_t$ and beyond that, we have no information about the distribution of $\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*$. Further, this non-asymptotic result provides no information on how the eigenvalues of $\boldsymbol{\Lambda}_t$ scale with t, leaving the dependence of the size of the confidence set on t unclear. We fill this gap within this paper, both by characterizing the asymptotic limiting distribution and by providing non-asymptotic guarantees for convergence to said limiting distribution. Unlike their confidence set, ours explicitly utilizes the quantile of the chi-squared distribution, entailing a stronger distributional result.

Regarding other known results, Theorem 3 in Lai and Wei (1982) establishes the asymptotic normality of the least squares estimator for the adaptive linear regression model. Here, in contrast to the multi-armed bandit setting, the "stability" condition requires that the sample covariance matrix (and not the sequence of arm pull rates) stabilizes to a deterministic sequence. They also demonstrate that asymptotic normality can fail in the absence of this stability property. In the multi-armed bandit (MAB) setting, a special case of linear bandits considered here, Khamaru and Zhang (2024) verifies the stability property and proves asymptotic normality for the UCB algorithm. In a broader context, Banerjee et al. (2023) controls the minimum eigenvalues of the design matrix generated by any linear bandit algorithm with sublinear regret. As we shall see shortly, this quantity also plays a central role in our analysis. However, it remains unclear whether the output of the LinUCB algorithm is asymptotically normal without additional assumptions, and if so, what the limiting covariance structure would be. This question is appealing both theoretically and practically: if asymptotic normality holds, one can construct substantially tighter confidence sets than those derived from concentration inequalities.

1.2 Our contributions

In this paper, we study the asymptotic behavior of the LinUCB algorithm in the setting where the true parameter lies on the unit sphere and the action set is the unit ball (Assumption 1), under sub-Gaussian noise (Assumption 2). We show that the algorithm is asymptotically normal, aligning with empirical results in Figure 1.
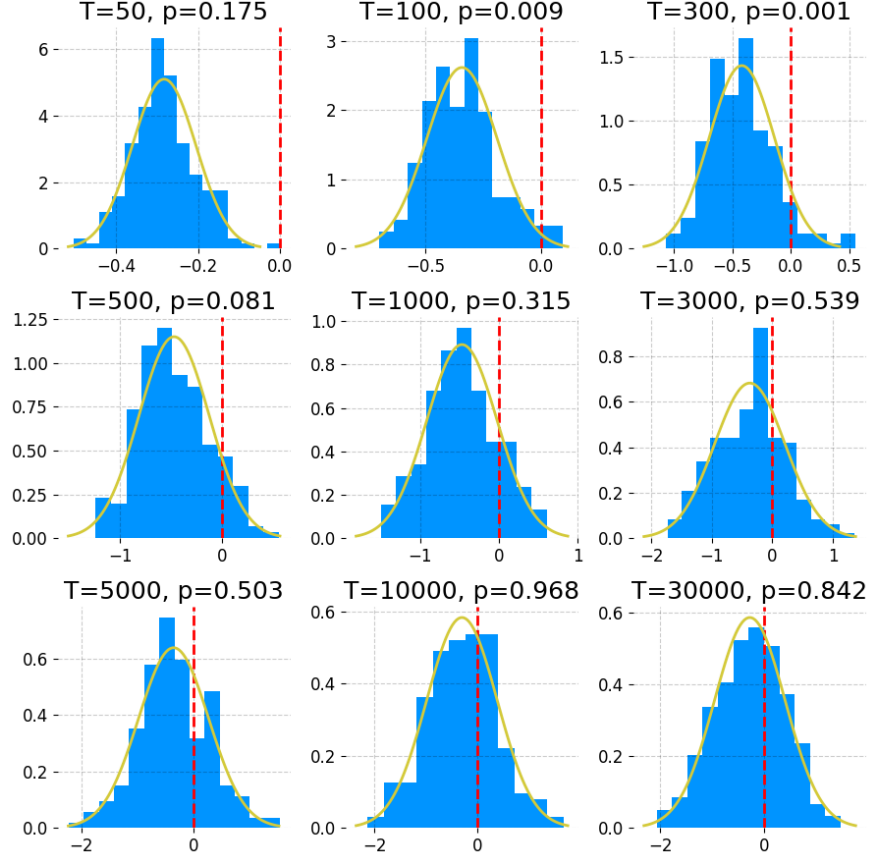
Figure 1: Asymptotic normality of the LinUCB algorithm in case where the action set is the unit ball. For some random vector u on the unit ball, we plot $\widehat{\sigma}^{-1}\left(\frac{2\beta^2 T}{d+1}\right)^{1/4}u^\top\left(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star\right)$ over 1000 independent trials, with KDE estimate overlaid as well as Shapiro-Wilk p-values provided as a test for non-normality. Asymptotic normality is indeed demonstrated, but the rate of convergence to the true parameter is certainly empirically slower than the $1/\sqrt{T}$ parametric rate, corroborating our theory.

- Our main result, Theorem 1 shows that, when projected onto $(\boldsymbol{\theta}^\star)^\perp$ (the orthogonal complement of $\boldsymbol{\theta}^\star$), the asymptotic covariance of $\widehat{\boldsymbol{\theta}}_T$ is isotropic—of magnitude $\sqrt{\frac{d+1}{2\beta^2 T}}\,\boldsymbol{I}_{d-1}$ for a broad class of exploration schedules $\beta = \beta(T, d)$. Equivalently, after an explicit rescaling proportional to β, the projected error obeys a central limit theorem: for any $\boldsymbol{U} \in \mathbb{R}^{d \times (d-1)}$ with orthonormal columns orthogonal to $\boldsymbol{\theta}^\star$,

$$\left(\frac{2\beta^2 T}{d+1}\right)^{1/4} \boldsymbol{U}^\top\left(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2 \boldsymbol{I}_{d-1}\right).$$

Since $\widehat{\boldsymbol{\theta}}_T$ lies on the unit sphere, this essentially pins down the full asymptotic law of $\widehat{\boldsymbol{\theta}}_T$. To our knowledge, these results give the first asymptotic convergence guarantee for the parameter estimate based on a dependent, adaptively collected data sequence generated by the LinUCB policy.

- The above result allows us to provide an asymptotic $(1-\delta)$ Wald-type confidence set for $\boldsymbol{\theta}^\star$. In contrast to the confidence set from Abbasi-yadkori et al. (2011) in (2), we require only an asymptotically spherical confidence set

$$\mathcal{C}_\delta = \left\{\boldsymbol{\theta} \in \mathcal{S}^{d-1} : \left\|\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}\right\|_2^2 \leq \widehat{\sigma}^2 \sqrt{\frac{d+1}{2\beta^2 T}} \cdot \chi^2_{d-1, 1-\delta}\right\},$$

for which we can provide a precise and deterministic (modulo randomness in the variance estimate $\widehat{\sigma}^2$) characterization of its diameter. If the user desires an ellipsoid confidence set for better finite-sample performance, they can utilize

$$\mathcal{C}_\delta^{\mathsf{ellipsoid}} = \left\{ \boldsymbol{\theta} \in \mathcal{S}^{d-1} : \left\| \widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta} \right\|_{\boldsymbol{\Lambda}_T}^2 \leq \widehat{\sigma}^2 \chi_{d-1,1-\delta}^2 \right\},$$

which is asymptotically equivalent to $\mathcal{C}_\delta$ while maintaining the same distributional guarantees. Whereas the LinUCB confidence set in (2) is derived via martingale concentration, our sets are standard Wald-type confidence sets used for asymptotically normal models under i.i.d. sampling. A key caveat is the rate: with T observations, the estimator concentrates at $O_p(T^{-1/4})$, in contrast to the $O_p(T^{-1/2})$ rate (and corresponding confidence-set width) in the i.i.d. setting. This perspective allows us to conduct statistical inference on data collected by LinUCB, paralleling the classical UCB algorithm for multi-armed bandits (Khamaru and Zhang, 2024).

- In the process of proving this result, we ended up proving several results that may be of independent interest. These include:

  - A tighter uniform control over the error of the estimated parameter at each timestep within Theorem 2. Instead of obtaining a high-probability uniform bound for $\|\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^\star\|_{\boldsymbol{\Lambda}_t}$ ($t \in [T]$) scaling in $O(\sqrt{\log T})$ as shown in (2), we obtain an improved bound of $O(\sqrt{\log \log T})$ when the failure probability is $1/\log T$.

  - A complete characterization of the eigenvalues of the feature covariance $\boldsymbol{\Lambda}_t$ under LinUCB. In particular, we describe how these eigenvalues evolve throughout the learning process, summarized in Propositions 1–4. First, in Propositions 1 and 2, we establish the early-stage pattern: all non-leading eigenvalues are of the same order – $\Theta(t)$ in the first phase and $\Theta(\sqrt{t})$ in the second. Next, Proposition 3 identifies the key transition, wherein the top eigenvector concentrates more tightly around $\boldsymbol{\theta}^\star$; building on this transition, Propositions 4 develop a fine-grained analysis of the non-leading eigenvalues and show that they converge to a deterministic limit.

Although analogous results are available for the multi-armed bandit setting with the UCB algorithm (Khamaru and Zhang, 2024), that setting features a finite, fixed action set. In contrast, our linear bandit setting allows arbitrary adaptive exploration over the unit ball, yielding a continuum of actions. This richer action space makes the asymptotic analysis substantially more delicate and requires techniques beyond those used in the finite-arm case.

1.3 Other related works

Bandit algorithms and inference for bandits. Dating back to the seminal works Robbins (1952); Thompson (1933), bandit algorithms have attracted tremendous attention for their simplicity and flexibility in modeling adaptive data collection. The UCB algorithm was first proposed in Lai and Robbins (1985); Lai (1987), and its linear extension, LinUCB, was introduced in Li et al. (2010). UCB and its many variants have since been widely applied to problems with dynamic data, leaving a profound impact across statistics, operations research, and reinforcement learning. Classical research has primarily focused on regret analysis, often establishing sublinear bounds for UCB and its extensions (see Bubeck et al. (2012); Lattimore and Szepesvári (2020) and references therein). More recent work has refined these results, providing precise regret bounds (Han et al. (2024); Fan et al. (2024)), and in some cases explicitly incorporating stability to enable tractable inference alongside learning (Sengupta and Khamaru (2024)). While the inferential properties of UCB have been recently studied in the multi-armed bandit setting Khamaru and Zhang (2024), much less is known about the inferential properties of LinUCB, beyond the non-asymptotic guarantees provided by confidence sets Abbasi-yadkori et al. (2011).

Inference with adaptively collected data. As discussed above, when data are collected adaptively, as in multi-armed and contextual bandits, standard i.i.d. asymptotics fail, since sampling depends on past observations, introduces complex dependencies (Lai and Wei (1982); Dwork et al. (2015)). To address

this challenge, a growing body of literature develops new statistical inference procedures that come with theoretical guarantees. Broadly, depending on the flavor of these results, they can be classified into two categories: those providing finite-sample guarantees and those establishing asymptotic characterizations. High probability bounds that hold for any finite samples often rely on tools such as concentration of self-normalized martingales (e.g. Abbasi-yadkori et al. (2011); Shi et al. (2023); Waudby-Smith et al. (2024); Nair and Janson (2023); Dimakopoulou et al. (2021); Wu et al. (2024)). In contrast, the asymptotic line of work focuses on characterizing the limiting distribution of estimators of interest, enabling confidence intervals and hypothesis testing in large-sample regimes (e.g. Hadad et al. (2021); Halder et al. (2025); Deshpande et al. (2018); Zhang et al. (2020); Niu and Ren (2025); Wu et al. (2025); Guo and Xu (2025)).

1.4 Organization and notation

Paper organization. The remainder of the paper is organized as follows. Section 2 reviews basics for linear bandits and the LinUCB algorithm, and introduces the stability condition necessary for establishing asymptotic results under LinUCB and other adaptive data-collection schemes. Section 3 presents our main results—an asymptotic normality theorem and an associated confidence set—along with further remarks and a proof sketch. Section 4 analyzes the evolution of the design covariance under LinUCB, a key ingredient in establishing our main result. Section 5 concludes our paper with a discussion and outlines several directions for future work.

Notation. Throughout this paper, $\|\boldsymbol{x}\|_2$ (or simply $\|\boldsymbol{x}\|$) denotes the Euclidean norm for a vector $\boldsymbol{x}$, and $\|\boldsymbol{A}\|_2$ (or $\|\boldsymbol{A}\|$) denotes the spectral (matrix 2-) norm for a matrix $\boldsymbol{A}$. For a positive definite matrix $\boldsymbol{\Lambda}$ and a vector $\boldsymbol{x}$, write the weighted norm as $\|\boldsymbol{x}\|_{\boldsymbol{\Lambda}} := \sqrt{\boldsymbol{x}^\top \boldsymbol{\Lambda} \boldsymbol{x}}$. Let $\mathcal{S}^{d-1} := \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\|_2 = 1\}$ be the unit sphere in $\mathbb{R}^d$ and $\mathcal{B}^d := \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\|_2 \leq 1\}$ be the unit ball in $\mathbb{R}^d$. Define the projection onto the unit sphere by $\mathcal{P}(\boldsymbol{x}) := \boldsymbol{x}/\|\boldsymbol{x}\|_2$ for $\boldsymbol{x} \neq \boldsymbol{0}$. For any vector $\boldsymbol{x} \in \mathbb{R}^d$, define the orthogonal complement of $\boldsymbol{x}$ as $\boldsymbol{x}^\perp = \{\boldsymbol{y} \in \mathbb{R}^d : \langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0\}$, a $(d-1)$ dimensional subspace of $\mathbb{R}^d$.

In addition, for any two functions $f(T)$ and $g(T)$, we write $f(T) \lesssim g(T)$ (equivalently, $f(T) = O(g(T))$) if there exists a constant $c_1 > 0$ such that $|f(T)| \leq c_1 |g(T)|$. Conversely, we denote $f(T) \gtrsim g(T)$ (equivalently, $f(T) = \Omega(g(T))$) if there exists a constant $c_2 > 0$ such that $|f(T)| \geq c_2 |g(T)|$. We also adopt $f(T) \asymp g(T)$ (equivalently, $f(T) = \Theta(g(T))$) to indicate that both $f(T) \lesssim g(T)$ and $f(T) \gtrsim g(T)$ hold simultaneously. We write $f(T) = O_p(g(T))$ if $f(T)/g(T)$ bounded in probability as $T \to \infty$. We write $f(T) = \widetilde{O}(g(T))$ to indicate that the $f(T) = O(g(T))$ holds up to logarithmic factors. Moreover, we denote $f(T) = o(g(T))$ if $f(T)/g(T) \to 0$ as $T \to \infty$, and we denote $f(T) \gg g(T)$ if $f(T)/g(T) \to \infty$ as $T \to \infty$. Finally, c and C denote universal constants that do not depend on T.

2 Preliminaries

2.1 Stochastic linear bandits

We formally define the stochastic linear bandits in d dimensions. This framework models sequential decision-making where the expected reward of each action is a linear function of its associated features. The learning procedure unfolds over a time horizon of T rounds. For clarity in what follows, we use the subscript T to denote terminal quantities (i.e., those evaluated after T rounds), and the subscript $t \in [T] := 1, \ldots, T$ to denote per-round quantities.

In each round $t = 1, 2, \ldots, T$, the learner observes a context $\boldsymbol{x}_t \in \mathcal{X}$, and is presented a finite or infinite action set $\mathcal{A}_t$. We denote by $\mathcal{X}$ the context space and by $\mathcal{A} \supseteq \bigcup_{t=1}^T \mathcal{A}_t$ the overall action space, i.e., the union of all actions available across rounds. Given context $\boldsymbol{x} \in \mathcal{X}$, and action $\boldsymbol{a} \in \mathcal{A}$, we assume access to a feature mapping $\boldsymbol{\phi} : \mathcal{X} \times \mathcal{A} \to \mathbb{R}^d$. We define

$$\boldsymbol{\Phi}_t = \left\{ \boldsymbol{\phi}(\boldsymbol{x}_t, \boldsymbol{a}_t) : \boldsymbol{a}_t \in \mathcal{A}_t \right\} \tag{3}$$

be the set of features available to the learner at time t.

As briefly introduced in (1), the learner plays an action $\boldsymbol{a}_t \in \mathcal{A}_t$ based on the trajectory of previous actions and rewards $\tau^{(t)} := \{\boldsymbol{a}_1, r_1, \ldots, \boldsymbol{a}_{t-1}, r_{t-1}\}$ and receives a reward

$$r_t = \langle \boldsymbol{\phi}(\boldsymbol{x}_t, \boldsymbol{a}_t), \boldsymbol{\theta}^\star \rangle + \epsilon_t,$$

where $\boldsymbol{\theta}^\star$ is an unknown parameter that defines the expected reward function. We assume without loss of generality that $\|\boldsymbol{\theta}^\star\|_2 = 1$. The expected reward $\langle \boldsymbol{\phi}(\boldsymbol{x}_t, \boldsymbol{a}_t), \boldsymbol{\theta}^\star \rangle$ is a linear function with respect to $\boldsymbol{\theta}^\star$ and ϵ_t is noise that satisfies $\mathbb{E}[\epsilon_t|\mathcal{F}_{t-1}] = 0$. The feature map can be a neural embedding, random Fourier feature map, polynomial embedding, kernel embedding, or other similar feature maps. All we shall require is that it is known that the features are bounded at every timestep, and that the rewards are a linear function of the features. As such, and especially when pretrained embeddings are readily available, the linear bandit model can be surprisingly expressive.

The learner's goal is to minimize the cumulative regret, which characterizes the difference between the total expected reward and the best possible reward that the learner could possibly obtained. Formally speaking, the regret after round T is defined as

$$R_T := \sum_{t=1}^{T} \langle \boldsymbol{\phi}(\boldsymbol{x}_t, \boldsymbol{a}_t^\star), \boldsymbol{\theta}^\star \rangle - \sum_{t=1}^{T} \langle \boldsymbol{\phi}(\boldsymbol{x}_t, \boldsymbol{a}_t), \boldsymbol{\theta}^\star \rangle, \tag{4}$$

where $\boldsymbol{a}_t^\star = \operatorname{argmax}_{\boldsymbol{a} \in \mathcal{A}_t} \langle \boldsymbol{\phi}(\boldsymbol{x}_t, \boldsymbol{a}), \boldsymbol{\theta}^\star \rangle$, is defined as the best possible action that the learner could take at time t. When the action sets $\mathcal{A}_1 = \ldots = \mathcal{A}_T = \mathcal{A}$ are the same for all $t = 1, \ldots, T$ and the context is the same for all t as well, the expression then simplifies:

$$R_T := \sum_{t=1}^{T} \langle \boldsymbol{\phi}(\boldsymbol{x}, \boldsymbol{a}^\star) - \boldsymbol{\phi}(\boldsymbol{x}, \boldsymbol{a}_t), \boldsymbol{\theta}^\star \rangle, \qquad \boldsymbol{a}^\star = \operatorname{argmax}_{\boldsymbol{a} \in \mathcal{A}} \langle \boldsymbol{\phi}(\boldsymbol{x}, \boldsymbol{a}), \boldsymbol{\theta}^\star \rangle. \tag{5}$$

2.2 The LinUCB algorithm

A widely used and conceptually elegant strategy for minimizing regret in linear bandits is the Upper Confidence Bound (UCB) principle, adapted to the linear setting as *LinUCB*. The method embodies *optimism in the face of uncertainty*: the learner maintains a high-probability confidence region for the unknown parameter $\boldsymbol{\theta}^\star \in \mathbb{R}^d$ and acts as if the most favorable parameter in this region were the truth. Concretely, at each round t the learner assigns to every candidate action $\boldsymbol{a} \in \mathcal{A}_t$ a UCB score that trades off predicted reward and an uncertainty bonus:

$$\mathrm{UCB}_t(\boldsymbol{a}) = \langle \boldsymbol{\phi}(\boldsymbol{x}_t, \boldsymbol{a}), \widehat{\boldsymbol{\theta}}_{t-1} \rangle + \beta \sqrt{\boldsymbol{\phi}(\boldsymbol{x}_t, \boldsymbol{a})^\top \boldsymbol{\Lambda}_{t-1}^{-1} \boldsymbol{\phi}(\boldsymbol{x}_t, \boldsymbol{a})}, \tag{6}$$

where the first term is the estimated reward and the second is a data-dependent exploration bonus measuring uncertainty along the direction $\boldsymbol{\phi}(\boldsymbol{x}_t, \boldsymbol{a})$. To rationalize this principle, we detail these two ingredients of the UCB score in (6).

Estimated reward. At time $t-1$, define the cumulative covariance matrix

$$\boldsymbol{\Lambda}_{t-1} = \lambda \boldsymbol{I}_d + \sum_{s=1}^{t-1} \boldsymbol{\phi}(\boldsymbol{x}_s, \boldsymbol{a}_s) \boldsymbol{\phi}(\boldsymbol{x}_s, \boldsymbol{a}_s)^\top. \tag{7}$$

The (ridge) regularized least-squares estimator is then

$$\overline{\boldsymbol{\theta}}_{t-1} \in \operatorname*{arg\,min}_{\boldsymbol{\theta} \in \mathbb{R}^d} \left\{ \sum_{s=1}^{t-1} (r_s - \langle \boldsymbol{\phi}(\boldsymbol{x}_s, \boldsymbol{a}_s), \boldsymbol{\theta} \rangle)^2 + \lambda \|\boldsymbol{\theta}\|_2^2 \right\} = \boldsymbol{\Lambda}_{t-1}^{-1} \sum_{s=1}^{t-1} \boldsymbol{\phi}(\boldsymbol{x}_s, \boldsymbol{a}_s) \, r_s, \tag{8}$$

Because the ground truth $\boldsymbol{\theta}^\star$ has unit norm, we project the ridge estimate onto the unit sphere $\mathcal{S}^{d-1}$:

$$\widehat{\boldsymbol{\theta}}_{t-1} = \mathcal{P}(\overline{\boldsymbol{\theta}}_{t-1}), \tag{9}$$

With $\widehat{\boldsymbol{\theta}}_{t-1}$ in hand, we score any candidate action $\boldsymbol{a}$ at time t by the estimated reward $\langle \boldsymbol{\phi}(\boldsymbol{x}_t, \boldsymbol{a}), \widehat{\boldsymbol{\theta}}_{t-1} \rangle$. In practice, it is common to set $\lambda = 1$ and $\boldsymbol{\Lambda}_0 = \boldsymbol{I}_d$, which keeps the cumulative covariance $\boldsymbol{\Lambda}_t$ invertible in the early stages while balancing bias and variance.

Algorithm 1 Linear UCB Algorithm

1: **Input:** Horizon T, action set $\mathcal{A}_t$ and feature map ϕ, and exploration bonus β.
2: **Initialize:** $\boldsymbol{\Lambda}_0 = \boldsymbol{I}_d$.
3: **for** each round $t = 1, 2, \ldots, T$ **do**
4:      Compute $\overline{\boldsymbol{\theta}}_{t-1}$ as in (8) and $\widehat{\boldsymbol{\theta}}_{t-1}$ as in (9).
5:      For each $\boldsymbol{a} \in \mathcal{A}_t$, compute UCB score $\mathrm{UCB}_t(\boldsymbol{a})$ as in (6).
6:      Select and play action: $\boldsymbol{a}_t = \arg\max_{\boldsymbol{a} \in \mathcal{A}_t} \mathrm{UCB}_t(\boldsymbol{a})$ and observe reward r_t.
7:      Update $\boldsymbol{\Lambda}_t = \boldsymbol{\Lambda}_{t-1} + \phi(\boldsymbol{x}_t, \boldsymbol{a}_t)\phi(\boldsymbol{x}_t, \boldsymbol{a}_t)^{\top}$.
8: **end for**

Exploration bonus. The bonus term is set to be the conventional choice $\beta\sqrt{\phi(\boldsymbol{x}_t, \boldsymbol{a})^{\top}\boldsymbol{\Lambda}_{t-1}^{-1}\phi(\boldsymbol{x}_t, \boldsymbol{a})}$. The scalar β is chosen so that, with probability at least $1 - \delta$, the expected reward of every action is upper-bounded by $\mathrm{UCB}_t(\boldsymbol{a})$ uniformly over t and $\boldsymbol{a}$ (the standard optimism property). For example, it is sufficient to choose β as (Abbasi-yadkori et al., 2011)

$$\beta = \sigma\sqrt{d\log\big(1 + TL^2/d\big) + 2\log(1/\delta)} + 1, \tag{10}$$

where σ is the sub-Gaussian parameter of the noise, and L is an upper bound for $\|\phi(\boldsymbol{x}_t, \boldsymbol{a}_t)\|_2$. With such a schedule, LinUCB achieves high-probability regret $\widetilde{O}(d\sqrt{T})$ when L is constant (Abbasi-yadkori et al., 2011).

Given the UCB score in (6), LinUCB selects at time t

$$\boldsymbol{a}_t \in \arg\max_{\boldsymbol{a} \in \mathcal{A}_t} \mathrm{UCB}_t(\boldsymbol{a}),$$

the action whose upper confidence bound on the expected reward is largest, thereby optimistically balancing exploitation of high estimated rewards with exploration of uncertain actions. The learner then observes the immediate reward r_t and update UCB score for the next round. The procedure is summarized in Algorithm 1.

2.3 Asymptotic normality of LinUCB with stability condition

Beyond minimizing cumulative regret, LinUCB also serves as a natural candidate for statistical inference purposes. LinUCB outputs an estimator $\widehat{\boldsymbol{\theta}}_T$ at time T via (9) with the sequence of feature vectors $\phi(\boldsymbol{x}_1, \boldsymbol{a}_1), \ldots, \phi(\boldsymbol{x}_T, \boldsymbol{a}_T)$ and rewards $r_1, \ldots, r_T$ collected adaptively. If one can characterize the limiting distribution of $\widehat{\boldsymbol{\theta}}_T$, Wald-type confidence sets can be constructed for $\boldsymbol{\theta}^{\star}$. In the fixed-design case, where the action sequence is chosen a priori, if the terminal design $\boldsymbol{\Lambda}_T$ satisfies $\lambda_{\min}(\boldsymbol{\Lambda}_T) \to \infty$ and the noise is i.i.d. with mean zero and variance σ^2, then the ridge estimator $\overline{\boldsymbol{\theta}}_T$ obeys $\boldsymbol{\Lambda}_T^{1/2}(\overline{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^{\star}) \to \mathcal{N}(0, \sigma^2\boldsymbol{I}_d)$ (see Example 2.28 in van der Vaart (2000)). The same asymptotic normality holds in the i.i.d. random-design setting, provided the design covariance matrix is full rank.

When samples are collected adaptively (e.g., using LinUCB), the situation is more subtle. The terminal design $\boldsymbol{\Lambda}_T$ is random and history-dependent; its spectrum can vary across runs, so a single deterministic normalization under which $\overline{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^{\star}$ has a normal limit may not exist; in particular, the usual Lindeberg–Feller CLT for deterministic designs is not directly applicable.

In prior work, additional regularity conditions are made to ensure asymptotic normality. One such condition is the so-called *stability* condition (Lai and Wei (1982)), which assumes that the cumulative covariance matrix $\{\boldsymbol{\Lambda}_T\}$ admits a deterministic limit:

Definition 1 (Stability). *The sequence of sample covariance matrices $\{\boldsymbol{\Lambda}_T\}$ is stable if there exists a sequence of deterministic positive definite matrices $\{\boldsymbol{\Sigma}_T\}$ such that $\boldsymbol{\Sigma}_T^{-1}\boldsymbol{\Lambda}_T \longrightarrow \boldsymbol{I}_d$.*

It is unclear that without the stability assumption or other similar assumptions, whether the asymptotic normality and, therefore statistical inference can be achieved.

3 Main results

In this section, we develop an asymptotic theory for LinUCB without imposing the aforementioned stability assumption. We begin by stating several mild assumptions about our model and a few notation in Section 3.1.

Our main results are provided in Section 3.2, followed by a few remarks and implications. We present the main proof strategies of Theorem 1 in Section 3.3.

3.1 Assumptions

In this work, we analyze a canonical and broadly applicable regime where at every round t, the learner may choose any vector inside the unit ball of $\mathbb{R}^d$ ($d \geq 2$) as the action. Formally,

Assumption 1 (Unconstrained action set on unit ball). *We assume that the set of feature maps is given by $\mathbf{\Phi}_t = \mathcal{B}^d$ ($d \geq 2$) for all $t \in [T]$, where $\mathcal{B}^d$ denotes the unit ball in $\mathbb{R}^d$.*

The unit-ball action set allows the richest possible exploration directions. This stands in sharp contrast to the multi-armed bandit setting (Khamaru and Zhang, 2024), where the learner is limited to exploring a finite number of pre-specified directions. Under Assumption 1, the feature map is time-invariant with image $\mathbf{\Phi}_t \equiv \mathcal{B}^d$, the unit ball in $\mathbb{R}^d$. Consequently, choosing $\boldsymbol{a}_t$ given $\boldsymbol{x}_t$ is equivalent to selecting a point in $\mathcal{B}^d$. We therefore take $\mathcal{B}^d$ as the action set and, by slight abuse of notation, we identify each action with its feature vector, and write $\boldsymbol{a}_t \equiv \boldsymbol{\phi}(\boldsymbol{x}_t, \boldsymbol{a}_t)$ throughout.

We further assume the noise sequence $\{\epsilon_t\}_{t=1}^T$ (as in (1)) is sub-Gaussian with parameter σ. More concretely, we assume:

Assumption 2 (Sub-Gaussian noise). *Let $(\mathcal{F}_t)_{t=0}^T$ denote the natural filtration generated by Algorithm 1; that is, $\mathcal{F}_t := \sigma(\{\boldsymbol{a}_1, r_1, \ldots, \boldsymbol{a}_t, r_t\})$. The noise $(\epsilon_t)_{t=1}^T$ is an $(\mathcal{F}_t)$-adapted martingale difference sequence with conditional mean zero and conditional variance σ^2, and is conditionally sub-Gaussian with variance proxy σ^2: for each $t \geq 1$ and all $\lambda \in \mathbb{R}$,*

$$\mathbb{E}[\epsilon_t | \mathcal{F}_{t-1}] = 0, \quad \mathrm{Var}[\epsilon_t | \mathcal{F}_{t-1}] = \sigma^2, \quad \mathbb{E}[\exp(\lambda \epsilon_t) \mid \mathcal{F}_{t-1}] \leq \exp\left(\frac{1}{2}\sigma^2\lambda^2\right).$$

The sub-Gaussian property ensures the concentration of self-normalized martingales, which we shall leverage to control the cumulative effect of noise under adaptive sampling (see Theorem 2). In fact, this assumption is crucial for establishing the asymptotic normality under adaptive sampling, whereas for non-adaptive algorithms, asymptotic normality often follows from much weaker conditions (e.g., finite moment conditions, etc.).

3.2 Main theorems: asymptotic normality and confidence set

We now state our main result on the asymptotic behavior of the LinUCB estimator $\widehat{\boldsymbol{\theta}}_T$. Since $\boldsymbol{\theta}^\star$ has unit norm, we project and obtain $\widehat{\boldsymbol{\theta}}_T$ on the unit sphere $\mathcal{S}^{d-1}$. Consequently, it cannot admit a nondegenerate d-dimensional limit in $\mathbb{R}^d$. Instead, its first-order fluctuations are confined to the orthogonal linear subspace of $\boldsymbol{\theta}^\star$, denoted as $(\boldsymbol{\theta}^\star)^\perp$. Accordingly, we characterize its limit distribution after projecting onto $(\boldsymbol{\theta}^\star)^\perp$.

Theorem 1 (Asymptotic normality for LinUCB). *Under Assumptions 1–2, fix any matrix $\boldsymbol{U} \in \mathbb{R}^{d \times (d-1)}$ with orthonormal columns orthogonal to $\boldsymbol{\theta}^\star$; equivalently, let $\boldsymbol{Q} = (\boldsymbol{\theta}^\star, \boldsymbol{U})$, then $\boldsymbol{Q}^\top \boldsymbol{Q} = \boldsymbol{I}_d$. With $\beta \gg d^2(\sigma\sqrt{d + \log\log T} + 1)$ and $\beta = O(\mathrm{poly}\log T)$, the estimator $\widehat{\boldsymbol{\theta}}_T$ in (9) satisfies Central Limit Theorem*

$$\left(\frac{2\beta^2 T}{d+1}\right)^{1/4} \boldsymbol{U}^\top \left(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star\right) \to \mathcal{N}(0, \sigma^2 \boldsymbol{I}_{d-1}), \qquad as \ T \to \infty. \tag{11}$$

In Theorem 1, the columns of matrix $\boldsymbol{U} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{d-1})$ form an orthonormal basis of $(\boldsymbol{\theta}^\star)^\perp$; hence

$$\boldsymbol{U}^\top \widehat{\boldsymbol{\theta}}_T = \left(\boldsymbol{u}_1^\top \widehat{\boldsymbol{\theta}}_T, \ldots, \boldsymbol{u}_{d-1}^\top \widehat{\boldsymbol{\theta}}_T\right)$$

gives the coordinates of the orthogonal projection of $\widehat{\boldsymbol{\theta}}_T$ onto $(\boldsymbol{\theta}^\star)^\perp$, namely $\boldsymbol{U}\boldsymbol{U}^\top \widehat{\boldsymbol{\theta}}_T$. Since the limiting covariance is $\sigma^2 \boldsymbol{I}_{d-1}$, this result is invariant to any choice of $\boldsymbol{U}$: the estimator converges at the same rate in every direction of $(\boldsymbol{\theta}^\star)^\perp$. While we state the result in terms of $\boldsymbol{U}^\top(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star)$, it is effectively an asymptotic

normality statement for $\widehat{\boldsymbol{\theta}}_T$ as well, since in a neighborhood of $\boldsymbol{\theta}^\star$ on $\mathcal{S}^{d-1}$, the mapping $\boldsymbol{\theta} \mapsto \boldsymbol{U}^\top \boldsymbol{\theta}$ gives a one-to-one (indeed, near isometric, see (12) below) local reparameterization.

Building on this result, we construct an asymptotic $(1-\delta)$ confidence set for $\boldsymbol{\theta}^\star$. A naïve attempt would invert the limit law of $\boldsymbol{U}^\top(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star)$, leading to a set defined by $\|\boldsymbol{U}^\top(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta})\|_2^2$; this is infeasible because $\boldsymbol{U}$ depends on the unknown $\boldsymbol{\theta}^\star$. Instead, we work directly with the Euclidean distance $\|\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}\|_2^2$. Using a local expansion around $\boldsymbol{\theta}^\star$, we have

$$\left\|\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star\right\|_2^2 = \left[1 + O\left(\left\|\boldsymbol{U}^\top(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star)\right\|_2^2\right)\right] \cdot \left\|\boldsymbol{U}^\top(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star)\right\|_2^2 = \left[1 + o_p(1)\right] \cdot \left\|\boldsymbol{U}^\top(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star)\right\|_2^2, \qquad (12)$$

where the last equality holds as $\|\boldsymbol{U}^\top(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star)\|_2^2 \to 0$, indicating these two criteria share the same asymptotic distribution. This allows us to construct the following confidence set of $\boldsymbol{\theta}^\star$.

Corollary 1 (Confidence set of LinUCB)**.** *Under Assumptions 1–2, when $\beta \gg d^2(\sigma\sqrt{d + \log\log T} + 1)$ and $\beta = O(\mathrm{poly}\log T)$, an asymptotic $(1-\delta)$ confidence set for $\boldsymbol{\theta}^\star$, based on $\widehat{\boldsymbol{\theta}}_T$, is given as*

$$\mathcal{C}_\delta = \left\{\boldsymbol{\theta} \in \mathcal{S}^{d-1} : \left\|\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}\right\|_2^2 \leq \widehat{\sigma}^2 \sqrt{\frac{d+1}{2\beta^2 T}} \cdot \chi^2_{d-1,1-\delta}\right\}, \qquad (13)$$

where $\widehat{\sigma}^2$ is the estimated noise variance, $\chi^2_{d-1,1-\delta}$ denotes the $(1-\delta)$-quantile of the χ^2 distribution with $d-1$ degrees of freedom.

Here, we note that the noise variance σ^2 can be consistently estimated due to the consistency of $\widehat{\boldsymbol{\theta}}_T$. Furthermore, the confidence set is spherical rather than not an ellipsoid, because, as shall be seen in Theorem 3, all non-leading eigenvalues converge to the same value asymptotically as $T \to \infty$. For moderate or small T, however, this convergence may be far from complete, so the empirical eigenvalues can still exhibit noticeable anisotropy. In such cases, the practitioner may prefer to capture the resulting ellipsoidal structure of the sampling distribution, as in the confidence set in (2). The following ellipsoidal confidence set provides an analogous asymptotic guarantee while accounting for this behavior:

$$\mathcal{C}_\delta^{\mathsf{ellipsoid}} = \left\{\boldsymbol{\theta} \in \mathcal{S}^{d-1} : \left\|\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}\right\|_{\boldsymbol{\Lambda}_T}^2 \leq \widehat{\sigma}^2 \chi^2_{d-1,1-\delta}\right\}. \qquad (14)$$

Next, we highlight several implications and consequences of our main results.

Convergence rate slowdown and effective sample size. With suitably chosen β, Theorem 1 together with Corollary 1 gives

$$\left\|\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star\right\|_2 = \widetilde{\Theta}_p\left(T^{-1/4}\right).$$

In contrast, the standard i.i.d. (parametric) rate is $\Theta_p(T^{-1/2})$. The slowdown arises because in the settings when the action set is fixed over time, regret-driven adaptivity concentrates actions near the optimum, making the terminal design covariance $\boldsymbol{\Lambda}_T$ ill-conditioned (its non-leading eigenvalues grow sublinearly in T), thereby yielding a slower convergence rate of the estimator $\widehat{\boldsymbol{\theta}}_T$.

To quantify this effect across adaptive data-collection regimes, we define the *effective sample size* as follows. If, for the terminal design covariance sequence $\{\boldsymbol{\Lambda}_T\}$, there exists a deterministic sequence $\{n_{\mathrm{eff},T}\}$ such that $n_{\mathrm{eff},T}^{-1}\lambda_{\min}(\boldsymbol{\Lambda}_T) \to 1$ with probability, then we refer to $n_{\mathrm{eff},T}$ as the effective sample size. The examples below show that $n_{\mathrm{eff},T}$ governs the estimator's convergence rate, yielding $\Theta_p\left(n_{\mathrm{eff},T}^{-1/2}\right)$.

- **Linear regression with i.i.d. design.** Consider the classical linear regression setting with i.i.d. design, in which the feature vector is sampled i.i.d. from some action set. Let $\boldsymbol{\Sigma} = \mathbb{E}[\boldsymbol{a}_1\boldsymbol{a}_1^\top]$ and assume $\boldsymbol{\Sigma}$ is full rank. It is easily seen that $\{\boldsymbol{\Lambda}_T\}$ is stable: with $\boldsymbol{\Sigma}_T^{-1}\boldsymbol{\Lambda}_T \to \boldsymbol{I}_d$ for $\boldsymbol{\Sigma}_T = T\boldsymbol{\Sigma}$. Hence the usual multivariate CLT applies: under standard regularity condition (e.g., mean-zero homoskedastic noise with variance σ^2),

$$\sqrt{T}\boldsymbol{\Sigma}^{1/2}(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star) \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I}_d).$$

  In this case, the effective sample size is $n_{\mathrm{eff},T} = T\lambda_{\min}(\boldsymbol{\Sigma}) \asymp T$, admitting the classical $T^{-1/2}$ convergence rate.

10

- **Multi-armed bandit with UCB.** Consider the classical K-armed bandit over T rounds. In our notation this corresponds to features $\boldsymbol{\phi}(\boldsymbol{x}_t, \boldsymbol{a}) = \boldsymbol{e}_a \in \mathbb{R}^K$ for arm a. At round t, using data up to time $t-1$, the UCB score for arm a is

$$\mathrm{UCB}_{a,t} = \overline{X}_{a,t-1} + \frac{\beta}{\sqrt{n_{a,t-1}}},$$

where $\overline{X}_{a,t-1}$ is the empirical mean reward of arm a and $n_{a,t-1}$ is the number of times arm a has been pulled up to $t-1$; $\beta > 0$ controls exploration. The algorithm then selects the arm with the largest score. Under mild regularity conditions, the allocation vector is asymptotically deterministic (Khamaru and Zhang, 2024; Han et al., 2024): there exist deterministic counts $\{n_{a,T}^\star\}_{a=1}^K$ such that $n_{a,T}/n_{a,T}^\star \to 1$ as $T \to \infty$. The limits are characterized by the "balancing" equations[1]

$$\mu_1 + \frac{\beta}{\sqrt{n_{1,T}^\star}} = \mu_2 + \frac{\beta}{\sqrt{n_{2,T}^\star}} = \cdots = \mu_K + \frac{\beta}{\sqrt{n_{K,T}^\star}}, \qquad \sum_{a=1}^K n_{a,T}^\star = T, \tag{15}$$

where μ_a is the mean reward of arm a. Then, for each arm,

$$\sqrt{n_{a,T}^\star}\left(\overline{X}_{a,T} - \mu_a\right) \xrightarrow{d} \mathcal{N}\!\left(0, \sigma_a^2\right),$$

where σ_a^2 is the reward variance for arm a. When all suboptimal arms have a fixed reward gap from the best arm, Khamaru and Zhang (2024) show that $n_{K,T}^\star = \Theta(\beta^2) = \Theta(\log T)$ in the limit. Equivalently, with the terminal design matrix $\boldsymbol{\Lambda}_T = \sum_{a=1}^K n_{a,T}\,\boldsymbol{e}_a \boldsymbol{e}_a^\top$ stabilizing to $\boldsymbol{\Sigma}_T = \sum_{a=1}^K n_{a,T}^\star\,\boldsymbol{e}_a \boldsymbol{e}_a^\top$, the effective sample size is $n_{\mathrm{eff},T} := \min_a n_{a,T}^\star = n_{K,T}^\star = \Theta(\log T)$. Consequently, the estimation error for suboptimal arms decays at rate $\Theta\!\left(1/\sqrt{\log T}\right) = \Theta\!\left(1/\sqrt{n_{\mathrm{eff},T}}\right)$.

Returning to our setting where the action set is the unit ball—a rich action set that covers the entire feature space—the effective sample size is

$$n_{\mathrm{eff},T} = \sqrt{\frac{2\beta^2 T}{d+1}} = \widetilde{\Theta}(\sqrt{T}),$$

Consequently, the estimator converges at a rate of inverse square-root of $n_{\mathrm{eff},T}$, i.e., $\widetilde{\Theta}_p(T^{-1/4})$. Relative to an i.i.d. design, the effective sample size is smaller for a fixed number of observations, leading to slower convergence. By contrast, it is substantially larger than in the multi-armed bandit UCB setting: with a continuous action set near the optimum, no single action dominates, so the learner continues to explore multiple directions, which increases regret ($\widetilde{\Theta}(\sqrt{T})$ for LinUCB versus $O(\sqrt{\log T})$ for UCB) yet yields better coverage (faster growth of $\lambda_{\min}(\boldsymbol{\Lambda}_T)$) and thus faster estimator convergence.

Comparison to the confidence set in Abbasi-yadkori et al. (2011). We compare the confidence set constructed in Corollary 1 with the confidence set in (2) from Abbasi-yadkori et al. (2011). Our confidence set utilizes the standard Wald-type construction, while that of Abbasi-yadkori et al. (2011) is based on martingale concentrations. Unlike their confidence set, ours explicitly utilizes the quantile of the chi-squared distribution instead of a $\log(1/\delta)$ concentration inequality-like term, thereby encompassing a stronger distributional result instead of only tail concentration.

Another key difference that we would like to point out is that our construction does not rely on the random, round-by-round empirical feature covariance accumulated by the algorithm – but can do so if the user desires an ellipsoid confidence set. As such, we can provide a precise and deterministic (modulo randomness in $\widehat{\sigma}^2$) characterization of its diameter. Our confidence set is asymptotically tighter by a factor of $\sqrt{\log T}$.

Most importantly, our confidence set is simply the Wald-type confidence set commonly employed within statistics for asymptotically normal models with i.i.d. data. As in Khamaru and Zhang (2024), this amounts to saying that we can treat the data collected by LinUCB when performing statistical inference as if it was i.i.d., just as with the UCB algorithm for multi-armed bandits.

---

[1] The form in (15) differs from Khamaru and Zhang (2024) but is equivalent.

3.3 Technical overview

We next point out several key steps in the proof of Theorem 1. Although the final result is of asymptotic flavor, the argument rests on a sequence of non-asymptotic results. In adaptive data collection, where each action depends on past observations, such finite-sample controls are crucial: they stabilize the (effective) design covariance, which is the key ingredient for the asymptotics. We now outline the main steps and the technical challenges of the argument.

Uniform bound on ridge estimation error. The first step of our proof is to establish a refined uniform bound for the ridge estimator when actions are selected by LinUCB. We first obtain a uniform $\mathbf{\Lambda}_t$-norm bound on the estimation error.

Theorem 2 (Uniform control of estimation error)**.** *With probability* $1 - \frac{1}{\log T}$, *the estimation error of ridge estimator* $\overline{\boldsymbol{\theta}}_t$ *scaled by cumulative covariance matrix* $\mathbf{\Lambda}_t$ *satisfies*

$$\max_{1 \leq t \leq T} \left\| \overline{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^\star \right\|_{\mathbf{\Lambda}_t} \lesssim \sigma \sqrt{d + \log \log T} + 1. \tag{16}$$

With the $\mathbf{\Lambda}_t$-norm bound in Theorem 2, we immediately obtain a Euclidean norm control:

Corollary 2. *With probability* $1 - \frac{1}{\log T}$, *the estimation error of ridge estimator* $\overline{\boldsymbol{\theta}}_t$ *can be uniformly upper bounded as*

$$\| \overline{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^\star \|_2 \lesssim \frac{\sigma \sqrt{d + \log \log T} + 1}{\sqrt{\lambda_{t,d}}}. \tag{17}$$

Furthermore, one has the following upper bound for the projected estimator $\widehat{\boldsymbol{\theta}}_T$:

$$\| \widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^\star \|_2 \lesssim \frac{\sigma \sqrt{d + \log \log T} + 1}{\sqrt{\lambda_{t,d}}}. \tag{18}$$

While a direct application of maximal concentration inequalities yields an upper bound of the estimation error of order $O(\sqrt{\log T})$, Theorem 2 and Corollary 2 establish a tighter $O(\sqrt{\log \log T})$ growth rate for the maximal normalized estimation error. This is achieved by leveraging the temporal correlation structure among the estimation errors. When the exploration parameter β is chosen to grow faster than $\log \log T$ (for instance, of order $\log T$ as in Abbasi-yadkori et al. (2011)), this result implies that the noise-induced estimation error is asymptotically dominated by the exploration bonus. This property constitutes a crucial component of our analysis, enabling a precise characterization of the asymptotic behavior of the UCB algorithm.

A related result was established in Lemma 5.1 of Khamaru and Zhang (2024),[2] though the analysis in our setting is considerably more involved due to the presence of a multi-dimensional noise term with a nonstationary sample covariance matrix. To control this term, we introduce an exponential supermartingale and control a weighted aggregation of it, before establishing uniform concentration over the aggregated process. This step requires constructing a net—chosen so that the maximum over the full region is effectively captured by its maximum on the net, while accommodating the process's nonstationary nature. In particular, we construct a global net that jointly covers all covariance matrices and noise terms up to time T, yet has only $O(\log T)$ cardinality. The construction leverages the rare-switching technique of Abbasi-yadkori et al. (2011), which is commonly used to control the growth of function classes in online reinforcement learning (He et al., 2023; Sherman et al., 2024; Tan et al., 2025). However, we employ this idea differently to build an $O(\log T)$-sized collection of representative time indices which forms a covering net such that the associated covariance matrices collectively approximate all covariance matrices. The detailed proof is provided in Appendix C.

Characterization of $\mathbf{\Lambda}_T$. In the next step, we provide a careful characterization of the eigenstructure of the terminal design $\mathbf{\Lambda}_T$, which plays a critical role in developing Theorem 1. The analysis hinges on Theorem 2, which shows that, as T grows, the noise has minimal influence compared with the exploration bonus, rendering the former asymptotically negligible. Theorem 3 formalizes the result, establishing (i) convergence of the top eigenvector and (ii) concentration of the non-leading eigenvalues around a deterministic limit.

---

[2]That result was derived using the argument presented in http://blog.wouterkoolen.info/QnD_LIL/post.html.
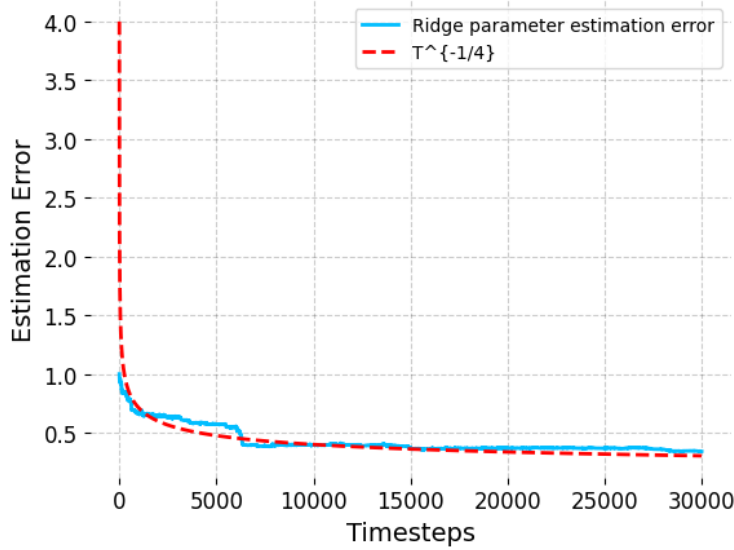
Figure 2: Estimation error of the parameter estimate obtained by ridge regression within the LinUCB algorithm. That is, we plot $\|\overline{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^\star\|_2$ against timesteps. In this simulation, the action set is the unit ball and the optimal parameter is the first standard basis vector. We see that the estimation error decreases according to the $T^{-1/4}$ rate as predicted within Theorem 1.

Theorem 3 (Eigenstructure concentration of LinUCB). *Under Assumptions 1–2, let $\{\lambda_{T,i}\}_{i=1}^d$ be the eigenvalues of $\boldsymbol{\Lambda}_T$ ordered non-increasingly $\lambda_{T,1} \geq \cdots \geq \lambda_{T,d}$, and let $\{\boldsymbol{v}_{T,i}\}_{i=1}^d$ be the corresponding eigenvectors. If $\beta \gg d^2(\sigma\sqrt{d + \log\log T} + 1)$, then with probability $1 - \frac{1}{\log T}$,*

- **Alignment of the top eigenvector.** *The leading eigenvector $\boldsymbol{v}_{T,1}$ concentrates to the signal $\boldsymbol{\theta}^\star$,*

$$\|\boldsymbol{v}_{T,1} - \boldsymbol{\theta}^\star\|_2 \lesssim \frac{\sigma\sqrt{d + \log\log T} + 1}{\sqrt{\lambda_{T,d}}}. \tag{19}$$

- **Concentration of non-leading eigenvalues.** *The non-leading eigenvalues concentrate uniformly to a deterministic limit: for any $i \geq 2$*

$$\lambda_{T,i} = \left[1 + O\left(d\left(\frac{\beta^8}{T\sigma^6}\right)^{\frac{d+1}{d-1}} + \left(\frac{\sigma\sqrt{d + \log\log T} + 1}{\beta}\right)^{1/2}\right)\right]\sqrt{\frac{2\beta^2 T}{d+1}}. \tag{20}$$

Specifically, when $\beta = O(\mathrm{poly}\log T)$, we have

$$\lambda_{T,i} = (1 + o(1))\sqrt{\frac{2\beta^2 T}{d+1}}. \tag{21}$$

The proof of Theorem 3 is the most technically involved part of our analysis. Our approach provides a precise account of how the cumulative covariance matrix $\boldsymbol{\Lambda}_t$ ($t \in [T]$) evolves through four distinct phases; each phase calls for a different analysis. We outline the key arguments by phase in Section 4 and present the full proof in Appendix D.

Stabilizing projected covariance and establishing a CLT. Lastly, we derive sharp characterizations for the *projected* design covariance. Since the asymptotic variance of the projected estimator $\widehat{\boldsymbol{\theta}}_T$ is controlled by $\widetilde{\boldsymbol{\Lambda}}_T^{-1} = (\boldsymbol{U}^\top \boldsymbol{\Lambda}_T \boldsymbol{U})^{-1}$, it is sufficient to analyze the stability of projected sequence $\{\widetilde{\boldsymbol{\Lambda}}_T\}$ rather than the full sequence $\{\boldsymbol{\Lambda}_T\}$. For the diagonal matrix sequence

$$\widetilde{\boldsymbol{\Sigma}}_T := \sqrt{\frac{2\beta^2 T}{d+1}}\,\boldsymbol{I}_{d-1},$$

13

we prove in Section B that with probability $1 - \frac{1}{\log T}$,

$$\big\|\widetilde{\boldsymbol{\Sigma}}_T^{-1}\widetilde{\boldsymbol{\Lambda}}_T - \boldsymbol{I}_{d-1}\big\|_2 \; \lesssim \; d\bigg(\frac{\beta^8}{T\sigma^6}\bigg)^{\frac{d+1}{d-1}} \; + \; \bigg(\frac{\sigma\sqrt{d + \log\log T} + 1}{\beta}\bigg)^{1/2}. \tag{22}$$

When $\beta \gg d^2(\sigma\sqrt{d + \log\log T} + 1)$ and $\beta = O(\text{poly}\log T)$, the right hand side of (22) vanishes, which suggests that $\{\widetilde{\boldsymbol{\Lambda}}_T\}$ and $\{\widetilde{\boldsymbol{\Sigma}}_T\}$ are asymptotically equivalent. Intuitively, β trades off exploration and stability: if β is too small, the policy chases noise, the design fails to stabilize, and the estimator cannot achieve clean asymptotic normality; on the other hand, if β is too large, the policy over-explores, dispersing samples and slowing the concentration of information toward the desired limit. The asymptotic normality of $\widehat{\boldsymbol{\theta}}_T$ then follows directly from CLT with Lyapunov condition. A complete derivation of this part appears in Appendix B.

4 Non-asymptotic evolution of cumulative covariance

We now present a precise, non-asymptotic characterization of the cumulative covariance matrix $\boldsymbol{\Lambda}_t$. Its evolution over t unfolds in four qualitatively distinct phases, governed by the changing influence of the bonus term; this structure, in turn, yields progressively sharper conclusions on the concentration of eigenvalues and eigenvectors as the cumulative covariance matrix $\boldsymbol{\Lambda}_t$ transitions across phases. We formalize each phase in Propositions 1–4, which together provide a phase-wise characterization of $\boldsymbol{\Lambda}_t$. Specializing to $t = T$, Theorem 3 follows directly from Propositions 3 and 4.

First, let us recall several notation. We use $\lambda_{t,1} \geq \cdots \geq \lambda_{t,d}$ to denote the eigenvalues ranked non-increasingly, and use $\overline{\lambda}_t$ as the average of the non-leading eigenvalues $\sum_{i=2}^{d} \lambda_{t,i}/(d-1)$. Denote the leading eigenvector as $\boldsymbol{v}_{t,1}$.

- **Phase I (Initial exploration of all directions).** From the start, the minimum eigenvalue $\lambda_{t,d}$ and $\overline{\lambda}_t$ grow at the same deterministic rate: there exists a deterministic sequence $\{\lambda_t^\star\}_{t \geq 0}$ such that $\lambda_{t,d} \asymp \overline{\lambda}_t \asymp \lambda_t^\star$ for all $t \geq 0$. In particular, during Phase I (i.e., for $t \leq t_1$), we have $\lambda_t^\star = \Theta(t)$.

- **Phase II ($\Theta(\sqrt{t})$ growth of non-leading eigenvalues).** For $t \geq t_1$, the comparability $\lambda_{t,d} \asymp \overline{\lambda}_t \asymp \lambda_t^\star$ persists, now with $\lambda_t^\star = \Theta(\sqrt{t})$. Meanwhile, the leading direction concentrates around $\boldsymbol{\theta}^\star$: uniformly over t, $\|\boldsymbol{v}_{t,1} - \boldsymbol{\theta}^\star\|_{\boldsymbol{\Lambda}_t} = O(\beta)$. Phase II continues until t_2, where a sharper bound for the leading eigenvector takes effect.

- **Phase III (Refined concentration of leading direction).** For $t \geq t_2$, the leading eigenvector $\boldsymbol{v}_{t,1}$ concentrates further around $\boldsymbol{\theta}^\star$: the weighted error satisfies $\|\boldsymbol{v}_{t,1} - \boldsymbol{\theta}^\star\|_{\boldsymbol{\Lambda}_t} = O(\sqrt{\log\log T})$, uniformly in t, improving upon the $O(\beta)$ bound from Phase II. Phase III lasts until t_3, when stronger control of the non-leading eigenvalues becomes available.

- **Phase IV (Concentration of non-leading eigenvalues).** For all $t \geq t_3$, we obtain the sharper concentration of non-leading eigenvalues: $\overline{\lambda}_t/\lambda_{t,d} = 1 + o(1)$, strengthening the earlier $O(1)$ comparability. In addition, both $\overline{\lambda}_t$ and $\lambda_{t,d}$ start to concentrate towards the deterministic sequence $\lambda_t^\star$; in particular, when $t = T$, $\lambda_{T,d} = (1 + o(1))\lambda_T^\star$ and $\overline{\lambda}_T = (1 + o(1))\lambda_T^\star$.

An illustration of the growth of the non-leading eigenvalues appears in Figure 3.

Phase I: Initial exploration of all directions. In the first phase, each direction is sampled only occasionally, so all eigenvalues of $\boldsymbol{\Lambda}_t$ are small. Because the action $\boldsymbol{a}_t$ is chosen by maximizing a UCB score based on $\boldsymbol{\Lambda}_{t-1}$, the confidence bonus $\beta \cdot (\boldsymbol{a}_t^\top \boldsymbol{\Lambda}_{t-1}^{-1} \boldsymbol{a}_t)^{1/2}$ dominates the predicted reward $\langle \boldsymbol{a}_t, \mathcal{P}(\widehat{\boldsymbol{\theta}}_{t-1}) \rangle$. This pushes actions toward the least-explored eigenspaces of $\boldsymbol{\Lambda}_{t-1}$. We formalize the resulting spectral growth below.

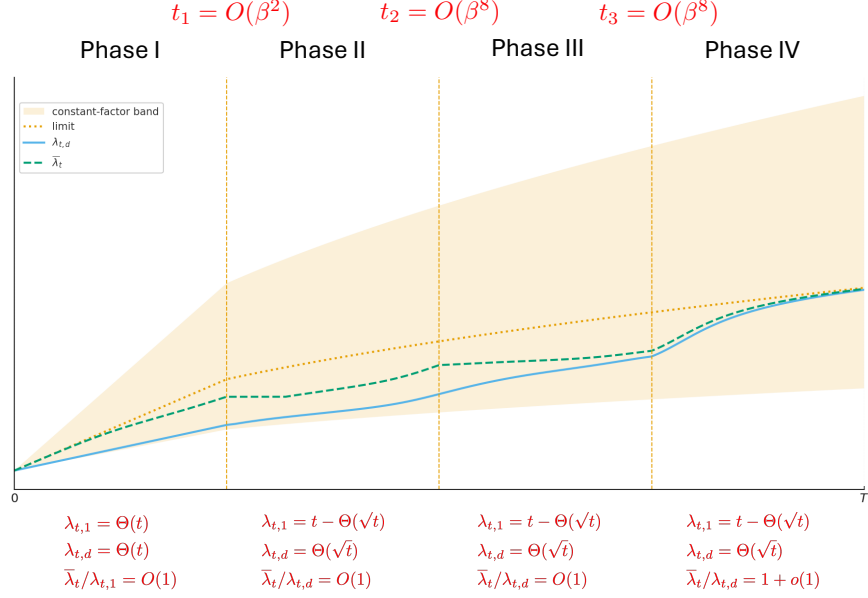**Proposition 1.** *For LinUCB (Algorithm 1), there exists $t_1 = \Theta(\beta^2 d)$ such that*

$$t_1 = O(\beta^2) \qquad t_2 = O(\beta^8) \qquad t_3 = O(\beta^8)$$

Phase I · Phase II · Phase III · Phase IV

$\lambda_{t,1} = \Theta(t)$	$\lambda_{t,1} = t - \Theta(\sqrt{t})$	$\lambda_{t,1} = t - \Theta(\sqrt{t})$	$\lambda_{t,1} = t - \Theta(\sqrt{t})$
$\lambda_{t,d} = \Theta(t)$	$\lambda_{t,d} = \Theta(\sqrt{t})$	$\lambda_{t,d} = \Theta(\sqrt{t})$	$\lambda_{t,d} = \Theta(\sqrt{t})$
$\overline{\lambda}_t / \lambda_{t,1} = O(1)$	$\overline{\lambda}_t / \lambda_{t,d} = O(1)$	$\overline{\lambda}_t / \lambda_{t,d} = O(1)$	$\overline{\lambda}_t / \lambda_{t,d} = 1 + o(1)$

Figure 3: Growth of $\lambda_{t,d}$ and $\overline{\lambda}_t$. Throughout the entire process, these two quantities grow on the same order, falling in a constant-factor band of a deterministic growth benchmark $\lambda_t^\star$. When $t \geq t_2$, the minimum eigenvalue $\lambda_{t,d}$ concentrates close to the non-leading mean $\overline{\lambda}_t$, and both $\lambda_{t,d}$ and $\overline{\lambda}_t$ concentrates to a deterministic limit $\lambda_t^\star$ when $t = T$.

- **(Linear growth of the minimum eigenvalue):** *For all $t \leq t_1$, the minimum eigenvalue $\lambda_{t,d}$ of $\mathbf{\Lambda}_t$ grows at least linearly, i.e.,*

$$\lambda_{t,d} \asymp \frac{t}{d}. \tag{23}$$

- **(Spectral gap between top two eigenvalues):** *With probability greater than $1 - 1/T$, the eigengap between the largest and the second largest eigenvalues at time t_1 is lower bounded as*

$$\lambda_{t_1,1} - \lambda_{t_1,2} \gtrsim t_1. \tag{24}$$

Proposition 1 certifies an initial *exploration-dominated* phase in which every action is explored approximately at the same rate, and eigenvalues grow roughly uniformly, that $\lambda_{t,d} \asymp t/d$. By time t_1, a pronounced gap separates the top two eigenvalues, signaling the onset of an *exploitation-dominated* stage in which one direction is revisited frequently. The scale $t_1 = \Theta(\beta^2 d)$ is natural: under $\lambda_{t,d} \asymp t/d$, the typical UCB width behaves as $\beta/\sqrt{\lambda_{t,d}} \asymp \beta/\sqrt{t/d}$ and becomes order one when $t \asymp \beta^2 d$. After that, the process transitions into the next phase.

Phase II: $O(\sqrt{t})$ growth of non-leading eigenvalues. In the subsequent phase, once the minimum eigenvalue $\lambda_{t,d}$ exceeds a fixed threshold, the estimator $\widehat{\boldsymbol{\theta}}_t$ attains a non-trivial correlation with the true signal $\boldsymbol{\theta}^\star$. From that point onward, action selection tilts toward the reward direction. Consequently, the top eigenvalue $\lambda_{t,1}$ absorbs most of the trace growth, while the non-leading eigenvalues grow with a slower diffusive rate. The precise growth rates and directional convergence are formalized next.

Proposition 2. *With probability greater than $1 - 1/T$, the following conditions hold for all $t_1 \leq t \leq T$.*

- *($O(\sqrt{t})$ **growth of the non-leading eigenvalues**) The non-leading eigenvalues grow with comparable speed. Specifically, the mean of non-leading eigenvalues $\overline{\lambda}_t$, and the minimum eigenvalue $\lambda_{t,d}$ satisfies*

$$\lambda_{t,d} \asymp \overline{\lambda}_t \asymp \beta \sqrt{\frac{t}{d}}. \tag{25}$$
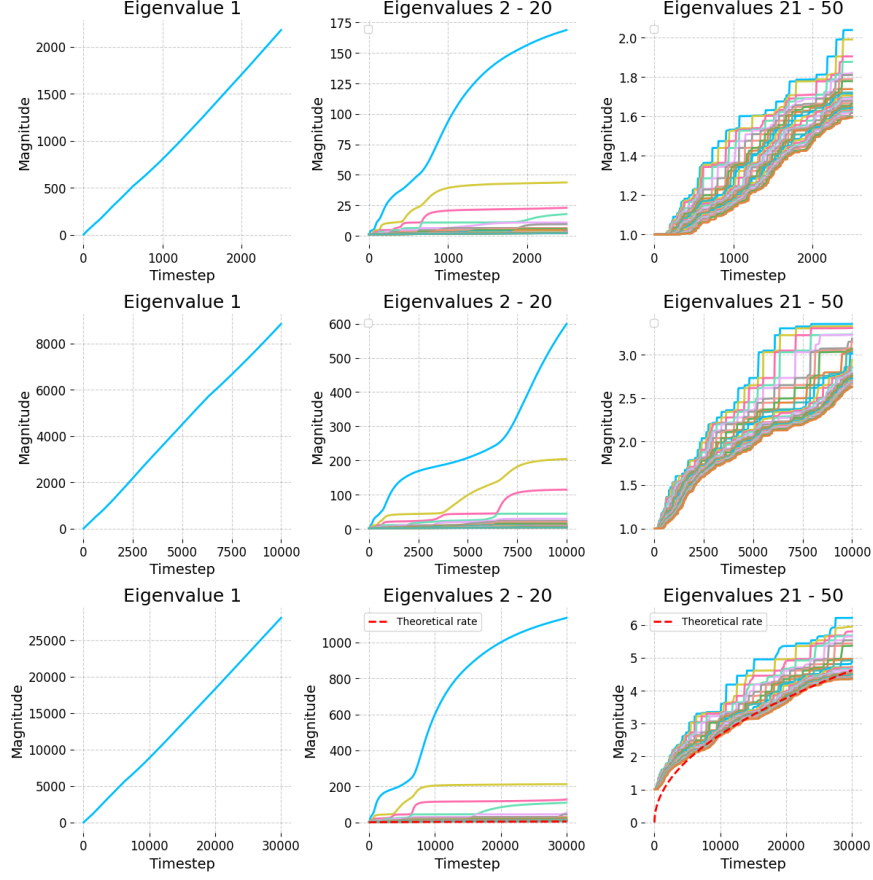
15

Figure 4: Rate of growth of the eigenvalues of the covariance matrix within a simulation of running LinUCB where the action set is the unit ball and the optimal parameter is the first standard basis vector. We see that the non-leading eigenvalues increase linearly at first as predicted by Proposition 1, before increasing on the order of $\sqrt{t}$ as predicted within Proposition 2. The dashed red line in the bottom row denotes the theoretical rate within Theorem 1. Overall, the above simulation aligns well with our theory.

- **(Concentration of leading eigenvector)** *The leading eigenvector* $\boldsymbol{v}_{t,1}$ *satisfies*

$$\|\boldsymbol{v}_{t,1} - \boldsymbol{\theta}^\star\|_2 \lesssim \frac{\beta}{\sqrt{\lambda_{t,d}}}. \tag{26}$$

The first part of this proposition guarantees that, beyond t_1, the non-leading spectrum is essentially flat and grows at the diffusive rate $\Theta(\sqrt{t})$: $\lambda_{t,d} \asymp \bar{\lambda}_t \asymp \beta\sqrt{t/d}$. Since each unit-norm action adds one to the trace, $\lambda_{t,1} = t - \Theta(\beta\sqrt{td})$, so $\lambda_{t,1} \asymp t$ and the top–second eigengap is linear in t. The second part controls the distance between the leading eigenvector with $\boldsymbol{\theta}^\star$ via standard eigenvector perturbation bounds. Notably, $\beta/\sqrt{\lambda_{t,d}}$ coincides with the worst-direction UCB width $\beta \cdot (\boldsymbol{a}^\top \boldsymbol{\Lambda}_t^{-1} \boldsymbol{a})^{1/2}$, so the leader's misalignment is controlled by the same quantity that governs exploration.

Phase III: Refined concentration of the top eigenvector. In this phase, we provide a sharper concentration guarantee for the leading eigenvector $\boldsymbol{v}_{t,1}$ of the cumulative covariance $\boldsymbol{\Lambda}_t$. Recall that, we have established a uniform high probability bound that $\|\boldsymbol{v}_{t,1} - \boldsymbol{\theta}^\star\|_2 = O(\beta/\sqrt{\lambda_{t,d}})$, matching the worst-direction exploration bonus, our goal for this stage is to obtain a finer result on this quantity.

Proposition 3. *There exists* $t_2 = O(\beta^8/(\sigma^6 d^2))$ *such that, with probability at least* $1 - \frac{1}{\log T}$*, the following holds simultaneously for* $t_3 \le t \le T$*: the leading eigenvector* $\boldsymbol{v}_{t,1}$ *satisfies*
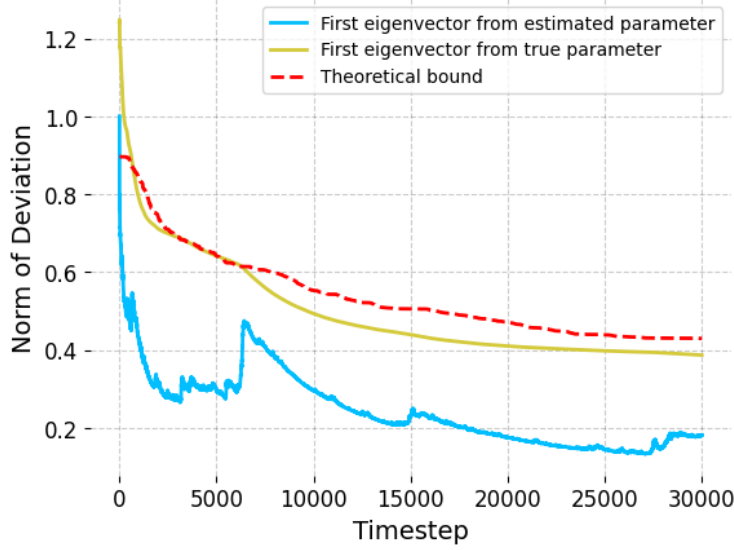
Figure 5: Concentration of the top eigenvector from the parameter estimate obtained through ridge regression and the true parameter. This is compared with the refined theoretical bound in Proposition 3.

$$\|\boldsymbol{v}_{t,1} - \boldsymbol{\theta}^\star\|_2 \lesssim \frac{\sigma\sqrt{d + \log\log T} + 1}{\sqrt{\lambda_{t,d}}}, \qquad \|\boldsymbol{v}_{t,1} - \widehat{\boldsymbol{\theta}}_t\|_2 \lesssim \frac{\sigma\sqrt{d + \log\log T} + 1}{\sqrt{\lambda_{t,d}}}. \qquad (27)$$

Proposition 3 shows that, for $t \geq t_2$, the misalignment of the leading direction $\|\boldsymbol{v}_{t,1} - \boldsymbol{\theta}^\star\|_2$ is of the same order as the maximum estimator error that shown in Corollary 2, both decaying like $(\sigma\sqrt{d + \log\log T} + 1)/\sqrt{\lambda_{t,d}}$. Since both the estimator $\widehat{\boldsymbol{\theta}}_t$ and the leading direction $\boldsymbol{v}_{t,1}$ concentrate tightly around the true signal $\boldsymbol{\theta}^\star$, they can be effectively treated as "quasi-deterministic" compared to the exploration bonus which scales as $O(\beta/\sqrt{\lambda_{t,d}})$. The evolution of non-leading eigenvalues, in the orthogonal space to $\boldsymbol{\theta}^\star$, depends mainly on the bonus term. This allows the non-leading eigenvalues to grow at comparable rates and concentrate accordingly—a mechanism that we formalize in the next phase.

Phase IV: Concentration of the non-leading eigenvalues. In this phase, we turn to the evolution of the *non-leading* eigenvalues of the cumulative covariance. Earlier phases established sharp control over the leading eigenvector, which capture the dominant direction of variability. Building on that, we now characterize the spectral structure on the subspace orthogonal to this top direction.

Proposition 4. *There exists $t_3 = O\left(\beta^8/\sigma^6\right)$ such that, with probability at least $1 - \frac{1}{\log T}$, the following holds simultaneously for all $t_3 \leq t \leq T$:*

- *(**Near equality of non-leading eigenvalues**) The non-leading eigenvalues concentrate as follows:*

$$\lambda_{t,2} = \left[1 + O\left(\frac{d(\sigma\sqrt{d + \log\log T} + 1)}{\beta}\right)\right]\lambda_{t,d}.$$

- *(**Deterministic benchmark and deviations**) The non-leading eigenvalues $\lambda_{t,i}$ $(i \geq 2)$ satisfies:*

$$\lambda_{t,i} = (1 + \Delta_{t,i})\sqrt{\frac{2\beta^2 t}{d+1}}, \qquad (28)$$

where $\Delta_{t,i}$ can be upper bounded as

$$|\Delta_{t,i}| \lesssim d\left(\frac{\beta^8}{t\sigma^6}\right)^{\frac{d+1}{d-1}} + \frac{(\sigma\sqrt{d + \log\log T} + 1)^{1/2}}{\sqrt{\beta}}, \qquad (29)$$

whenever $\beta \gtrsim d^2(\sigma\sqrt{d + \log\log T} + 1)$.

In words, Proposition 4 ensures that once $t \geq t_3$, the non-leading eigenvalues $\lambda_{t,2}, \ldots, \lambda_{t,d}$ are nearly equal and begin to concentrate around a deterministic value $\left(2\beta^2 t/(d+1)\right)^{1/2}$. More specifically, they satisfy

$$\frac{\lambda_{t,2}}{\lambda_{t,d}} = 1 + o(1),$$

so the associated eigenspace is approximately *isotropic*: restricted to the subspace orthogonal to the top eigenvector, the covariance is close to a scalar multiple of the identity. This near-isotropy indicates that, beyond the dominant signal direction, LinUCB explores the remaining directions at a nearly uniform rate. As a result, in the end when $t = T$, non-leading eigenvalues satisfy

$$\lambda_{T,i} = (1 + \Delta_{T,i})\left(\frac{2\beta^2 T}{d+1}\right)^{1/2}, \qquad 2 \leq i \leq d,$$

with $\Delta_{T,i}$ controlled by (29). This makes explicit the $\sqrt{t}$-scaling of the non-leading eigenvalues in directions orthogonal to the signal: the factor β^2 reflects the choice of LinUCB exploration bonus (which inflates uncertainty to promote exploration), while the normalization $2/(d+1)$ captures how the non-leading eigenvalues scale with dimension—when d increases, exploration is spread across more dimensions, and consequently each non-leading eigenvalue decreases.

5   Conclusion and future work

In this work, we characterize the asymptotic behavior of the LinUCB algorithm, and in doing so derive inference procedures that remain valid under adaptivity. Our main result shows the asymptotic normality of the terminal estimator of $\boldsymbol{\theta}^\star$: after an explicit rescaling by $\left(2\beta^2 T/(d+1)\right)^{1/4}$, the estimation error projected onto the tangent space satisfies a central limit theorem with variance $\sigma^2 \boldsymbol{I}_{d-1}$. This is accomplished through a thorough non-asymptotic characterization of the asymptotic behavior of the LinUCB feature covariance matrix, showing that it decomposes into a rank-one direction aligning with the true parameter and an isotropic bulk growing at a $\sqrt{T}$ rate.

However, our results pertain to the case where the action set is the unit ball. This is partly by design – rich action sets like these allow the learner to achieve good coverage over the feature space, but crucially, the learner does not need to (and in fact cannot) do so in order to achieve sublinear regret. With other choices of action sets, one should be able to achieve spiritually similar results, but the exact result will very much depend on whether these action sets ensure good coverage. For instance, within the MovieLens experiment within Kausik et al. (2024), a finite number of possible movies to recommend are sampled at each round, and the learner has to recommend the best one available. Here, as demonstrated in Figure 6, each possible action has good coverage on average, the minimum eigenvalue scales quickly on the order of T, and asymptotic normality is quickly achieved. We aim to explore this phenomenon in future work.

In addition, Berry-Esseen bounds characterizing distributional rates of convergence to the asymptotic distribution, as well as extensions to nonlinear function approximation methods and reinforcement learning (Wu et al., 2024, 2025), would also be welcome future directions to explore.

Ackowledgement

A   Technical preparations

We begin with the technical preliminaries needed for the proofs of the main theorems. Appendix A.1 details notation on the spectral decomposition of the design covariance, and Appendix A.2 lists auxiliary technical lemmas.
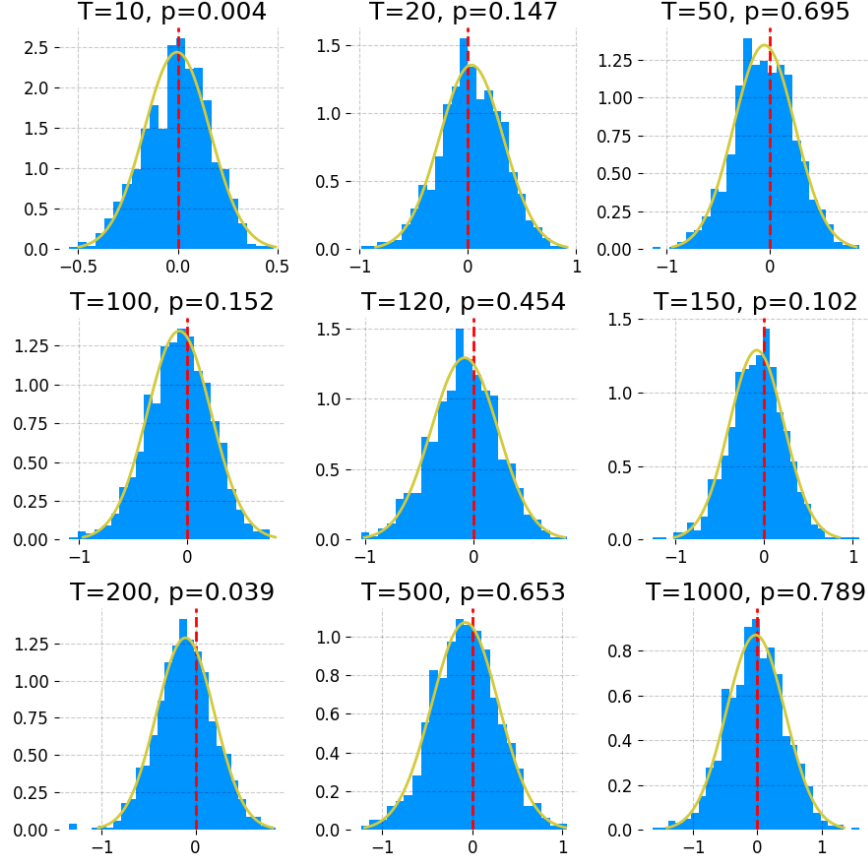
Figure 6: Asymptotic normality of the LinUCB algorithm in the setup of the MovieLens experiment within Kausik et al. (2024). For some random vector $\boldsymbol{u}$ on the unit ball, we plot $\sqrt{T} \cdot \boldsymbol{u}^\top(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star)$ over 1000 independent trials, with KDE estimate overlaid as well as Shapiro-Wilk p-values provided as a test for non-normality. A finite number of possible movies to recommend are sampled at each round, and the learner has to recommend the best one available. As each possible action has good coverage on average, the minimum eigenvalue scales quickly on the order of T, and asymptotic normality is quickly achieved.

A.1   Spectral decompositions

Throughout the proof, to analyze the spectral evolution of the design covariance matrix $\boldsymbol{\Lambda}_t$, we decompose $\boldsymbol{\Lambda}_t$ into its eigenvalues and eigenvectors. Moreover, we also decompose decompose the action vector $\boldsymbol{a}_t$ and the estimated signal $\widehat{\boldsymbol{\theta}}_t$ onto the orthogonal basis formed by the eigenvectors of $\boldsymbol{\Lambda}_t$.

Spectral decomposition $\boldsymbol{\Lambda}_t$.   As covariance matrix $\boldsymbol{\Lambda}_t$ is symmetric and positive semi-definite, it admits a spectral (or eigenvalue) decomposition of the form:

$$\boldsymbol{\Lambda}_t = \sum_{i=1}^{d} \lambda_{t,i} \boldsymbol{v}_{t,i} \boldsymbol{v}_{t,i}^\top, \tag{30}$$

where $\lambda_{t,1} \geq \lambda_{t,2} \geq \cdots \geq \lambda_{t,d}$ are the eigenvalues of $\boldsymbol{\Lambda}_t$ arranged in non-increasing order, and $\boldsymbol{v}_{t,1}, \ldots, \boldsymbol{v}_{t,d} \in \mathbb{R}^d$ are the corresponding eigenvectors. These eigenvectors satisfy the orthogonality condition $\boldsymbol{v}_{t,i}^\top \boldsymbol{v}_{t,j} = \delta_{ij}$, where δ_{ij} is the Kronecker delta. As a result, the collection $\{\boldsymbol{v}_{t,1}, \ldots, \boldsymbol{v}_{t,d}\}$ forms an orthonormal basis of $\mathbb{R}^d$, aligned with the principal directions of the covariance structure at time t.

Decomposition of a_t and $\widehat{\theta}_t$. To facilitate component-wise analysis, we further express the vectors a_t, $\widehat{\theta}_t$ in terms of the eigenbasis $\{v_{t,1}, \ldots, v_{t,d}\}$ derived from the spectral decomposition of Λ_t. That is,

$$a_t = \sum_{i=1}^{d} \kappa_{t,i} v_{t,i}, \quad \widehat{\theta}_t = \sum_{i=1}^{d} \nu_{t,i} v_{t,i}, \tag{31}$$

where the coefficients $\kappa_{t,i}, \nu_{t,i} \in \mathbb{R}$ represent the projections of a_t, $\widehat{\theta}_t$ onto the i-th eigenvector $v_{t,i}$, respectively. These coefficients are explicitly given by the inner products $\kappa_{t,i} = v_{t,i}^\top a_t$, $\nu_{t,i} = v_{t,i}^\top \widehat{\theta}_t$. With this decomposition, throughout the analysis, the vectors a_t and $\widehat{\theta}_t$ can be expressed in terms of their coordinates with respect to the orthonormal basis $\{v_{t,1}, \ldots, v_{t,d}\}$.

We further decompose the action vector a_t into its component along the estimated signal and its orthogonal complement:

$$a_t = \alpha_t \widehat{\theta}_t + \xi_t, \tag{32}$$

where $\widehat{\theta}_t$ is the ridge estimator projected onto the unit sphere, $\xi_t \in \mathbb{R}^d$ satisfies $\xi_t^\top \widehat{\theta}_t = 0$, and $\alpha_t \in \mathbb{R}$ is a scalar. Under Assumption 1, the LinUCB maximizer over the unit ball lies on the unit sphere—indeed, the radial projection $\mathcal{P}(a)$ of any interior point weakly increases $\mathrm{UCB}_t(a)$ defined in (6)—so $\|a_t\|_2 = 1$, which implies $\alpha_t^2 + \|\xi_t\|^2 = 1$. Intuitively, α_t measures alignment with the estimated signal, while ξ_t collects directions orthogonal to $\widehat{\theta}_t$. For fixed $(a_t, \widehat{\theta}_t)$, this orthogonal decomposition is unique.

A.2 Auxiliary lemmas

We now establish a series of technical lemmas that will be employed in the proof of our results. The proofs of the lemmas stated in this section are deferred to Appendix E.

Characterizing action vector a_t. Equipped with the spectral decompositions from Section A.1, we characterize the LinUCB action a_t in the orthonormal basis $\{v_{t,1}, \ldots, v_{t,d}\}$, which is key to tracking the evolution of the design covariance Λ_t. Our goal is to obtain a closed-form description of the coefficients $\kappa_{t,i}$. We begin with an equivalent representation of a_t.

Lemma 1 (An equivalent representation of a_t)**.** *The LinUCB action admits the following representation:*

$$w_t = \arg \max_{\|w\|_2 = 1} \left\| \widehat{\theta}_t + \beta \cdot \Lambda_t^{-1/2} w \right\|_2, \tag{33}$$

$$a_t = \mathcal{P}\left(\widehat{\theta}_t + \beta \Lambda_t^{-1/2} w_t \right), \tag{34}$$

where $\mathcal{P} : \mathbb{R}^d \to \mathcal{S}^{d-1}$ denotes the projection onto the unit sphere $\mathcal{S}^{d-1}$.

Lemma 1 shows that a_t is the projection of the sum of the current parameter estimate and an exploration shift. The shift $\beta \Lambda_t^{-1/2} w_t$ points toward under-explored (high-variance) directions, enabling a recursive description of the selected actions in terms of the projected ridge estimator $\widehat{\theta}_t$ in (9). Consequently, the coefficients $\kappa_{t,i}$ admit a closed form. Writing $w_t = (w_{t,1}, \ldots, w_{t,d})$, we have the equivalent optimization

$$w_t = \arg \max_{\|w\|_2 = 1} \sum_{i=1}^{d} \left(\nu_{t,i} + \frac{\beta w_i}{\sqrt{\lambda_{t,i}}} \right)^2, \quad \text{where } \sum_{i=1}^{d} \nu_{t,i}^2 = 1, \tag{35}$$

and the coefficients $\kappa_{t,i}$ are given by

$$\kappa_{t,i} = \frac{\nu_{t,i} + \frac{\beta w_{t,i}}{\sqrt{\lambda_{t,i}}}}{\sqrt{\sum_{j=1}^{d} \left(\nu_{t,j} + \frac{\beta w_{t,j}}{\sqrt{\lambda_{t,j}}} \right)^2}}. \tag{36}$$

This, in turn, enables us to quantify how the current action a_t decomposes into the component aligned with the estimated signal $\widehat{\theta}_t$ and the orthogonal component ξ_t, which is also related to the $\lambda_{t,d}$, the minimum eigenvalue of Λ_t.

Lemma 2 (Spectral decomposition of $\boldsymbol{a}_t$). *Suppose that*

$$\|\boldsymbol{v}_{t,1} - \widehat{\boldsymbol{\theta}}_t\|_2 \leq h_t, \tag{37}$$

the spectral decompositions in (31) and (32) satisfies

- *The decomposition of $\widehat{\boldsymbol{\theta}}_t$ on orthogonal basis formed by eigenvectors $\{\boldsymbol{v}_{t,1}, \ldots, \boldsymbol{v}_{t,d}\}$ satisfies:*

$$\nu_{t,1} \geq 1 - h_t^2, \quad \nu_{t,i} \leq h_t. \tag{38}$$

- *The decomposition of $\boldsymbol{a}_t$ on $\widehat{\boldsymbol{\theta}}_t$ satisfies:*

$$\alpha_t = 1 - O\left(\frac{\beta^2}{\lambda_{t,d}}\right), \quad \|\boldsymbol{\xi}_t\| = O\left(\frac{\beta}{\sqrt{\lambda_{t,d}}}\right). \tag{39}$$

- *The decomposition of $\boldsymbol{a}_t$ on orthogonal basis formed by eigenvectors $\{\boldsymbol{v}_{t,1}, \ldots, \boldsymbol{v}_{t,d}\}$ satisfies:*

$$\kappa_{t,1} = 1 - O\left(h_t^2 + \frac{\beta^2}{\lambda_{t,d}}\right), \quad \kappa_{t,i} = O\left(h_t + \frac{\beta}{\sqrt{\lambda_{t,d}}}\right), \text{ for any } i \geq 2. \tag{40}$$

In other words, we provide a quantitative characterization of the intuition that if the estimator $\widehat{\boldsymbol{\theta}}_t$ aligns well with $\boldsymbol{v}_{t,1}$, the leading eigenvector of $\boldsymbol{\Lambda}_t$, then the action $\boldsymbol{a}_t$ aligns well with both $\widehat{\boldsymbol{\theta}}_t$ and $\boldsymbol{v}_{t,1}$.

Fine grained characterization of $\boldsymbol{a}_t$. The lemmas above provide a coarse characterization of $\boldsymbol{a}_t$. While these formulas are clean and simple, they are not sufficient to establish finer properties of $\boldsymbol{a}_t$—in particular, to explain why $\boldsymbol{a}_t$ drives the non-leading eigenvalues of $\boldsymbol{\Lambda}_t$ to concentrate. To address this, we recast (35) as a constrained optimization problem. Fix a radius $c_0 > 0$, a scalar $\beta \in \mathbb{R}$, a signal vector $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_d) \in \mathbb{R}^d$, and a spectrum $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_d)$ with $\lambda_i > 0$ for all i. Define the objective

$$g(\boldsymbol{w}) := \sum_{i=1}^d \left(\nu_i + \frac{\beta w_i}{\sqrt{\lambda_i}}\right)^2,$$

and the maximizer over the ℓ_2–sphere of radius c_0,

$$\boldsymbol{w}^\star(c_0, \boldsymbol{\nu}, \boldsymbol{\lambda}) := \arg\max_{\|\boldsymbol{w}\|_2 = c_0} g(\boldsymbol{w}). \tag{41}$$

For bookkeeping, we also define the (scaled) coordinate contributions at the maximizer,

$$\kappa_i^\star(c_0, \boldsymbol{\nu}, \boldsymbol{\lambda}) := \frac{\left(\nu_i + \frac{\beta w_i^\star}{\sqrt{\lambda_i}}\right)^2}{\sqrt{\sum_{j=1}^d \left(\nu_j + \frac{\beta w_j^\star}{\sqrt{\lambda_j}}\right)^2}} = \frac{\left(\nu_i + \frac{\beta w_i^\star}{\sqrt{\lambda_i}}\right)^2}{\sqrt{g(\boldsymbol{w}^\star)}}. \tag{42}$$

Our first result establishes a lower bound on the projection of $\boldsymbol{a}_t$ onto the eigenspace associated with the "small eigenvalues". To avoid the extreme case where the optimization is driven solely by the signal vector $\boldsymbol{\nu}$, we impose the following structural condition.

Assumption 3. *There exists a constant $c > 0$ such that*

$$\max_{1 \leq i \leq d} \frac{\beta^2}{\lambda_i} \geq \frac{c}{c_0} \|\boldsymbol{\nu}\|_2^2.$$

Equivalently, $\beta^2/\lambda_{\min} \geq (c/c_0)\|\boldsymbol{\nu}\|_2^2$, so at least one rescaled coordinate (governed by β^2/λ_i) can compete with the signal energy $\|\boldsymbol{\nu}\|_2^2$. Hence the optimizer is influenced by the eigenstructure $\{\lambda_i\}$ rather than aligning with $\boldsymbol{\nu}$ alone. We then define for a fixed constant $c_1 > 1$ the index set of relatively small eigenvalues

$$\mathcal{L} := \left\{i : \lambda_i \leq c_1 \min_{1 \leq j \leq d} \lambda_j\right\}.$$

Then we have the following concentration property.

Lemma 3. *For $\boldsymbol{\kappa}^\star$ defined in (42), under Assumption 3, there exists a constant $C = C(c, c_1) > 0$ such that*

$$\sum_{i \in \mathcal{L}} \left(\kappa_i^\star \right)^2 \geq C \cdot c_0^2.$$

This "spectral concentration" means the optimal solution cannot spread its scaled mass arbitrarily; it must allocate a non-negligible portion to indices with small λ_i, where the factor $\beta/\sqrt{\lambda_i}$ enhances the coordinate-wise effect. The optimization balances two forces: the signal $\boldsymbol{\nu}$ and the spectral scaling $\beta/\sqrt{\lambda_i}$. Assumption 3 ensures the latter is sufficiently strong, and Lemma 3 shows the optimizer reflects this by concentrating on a subset of small-λ coordinates. We will use this later to control early-stage eigenvalue growth, in particular to show that certain eigenvalue ratios remain uniformly bounded under Assumption 3.

We then show another result related to the constrained optimization problem (41). Consider the modified problem with the canonical signal $\tilde{\boldsymbol{\nu}} = (1, 0, \ldots, 0)$. Then the solution $\boldsymbol{w}^\star$ is defined as

$$\widetilde{\boldsymbol{w}}^\star(c_0, \boldsymbol{\lambda}) := \boldsymbol{w}^\star(c_0, c_0\tilde{\boldsymbol{\nu}}, \boldsymbol{\lambda}) = \arg \max_{\|\boldsymbol{w}\|_2 = c_0} \left(c_0 + \frac{\beta w_1}{\sqrt{\lambda_1}} \right)^2 + \sum_{i=2}^d \frac{\beta^2 w_i^2}{\lambda_i}. \tag{43}$$

We compare the first coordinate of the optimizer in the general case to this canonical instance.

Lemma 4. *Let $\boldsymbol{\nu} \in \mathbb{R}^d$ satisfy $\|\boldsymbol{\nu}\|_2 = c_0$ and set $\boldsymbol{w}^\star(c_0, \boldsymbol{\nu}, \boldsymbol{\lambda})$ as in (41), and $\widetilde{\boldsymbol{w}}^\star(c_0, \boldsymbol{\lambda})$ as in (43). Then*

$$\left| w_1^\star(c_0, \boldsymbol{\nu}, \boldsymbol{\lambda}) \right| \leq \left| \widetilde{w}_1^\star(c_0, \boldsymbol{\lambda}) \right|.$$

This reduction is particularly convenient when translating coordinate bounds into statements about normalized contributions. For instance, any bound on $|\widetilde{w}_1^\star|$ immediately limits how much the term $\left(\nu_1 + \beta w_1^\star/\sqrt{\lambda_1} \right)^2$ can dominate the objective, and hence lower bound the contributions of $\boldsymbol{a}_t$ on non-leading eigenvalue directions. In later sections, we will exploit this "canonical-to-general" transfer to lower bound the growth of non-leading eigenvalues.

Rank-one update of $\boldsymbol{\Lambda}_t$.   To relate the action vector $\boldsymbol{a}_t$ to the evolution of the design covariance $\boldsymbol{\Lambda}_t$, we present the following lemma, which characterizes how the eigenvalues and eigenvectors of a positive-definite matrix evolve under a rank-one perturbation.

Lemma 5 (Theorem 8.4.3, Golub and Van Loan (2013))**.** *Let $\boldsymbol{A} \in \mathbb{R}^{d \times d}$ be a positive definite matrix with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d > 0$, and corresponding eigenvectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_d$. Let $\boldsymbol{u} = \sum_{i=1}^d \alpha_i \boldsymbol{v}_i$ for some scalars $\alpha_1, \ldots, \alpha_d \in \mathbb{R}$. Define the rank-one updated matrix $\widetilde{\boldsymbol{A}} = \boldsymbol{A} + \boldsymbol{u}\boldsymbol{u}^\top$.*

1. The eigenvalues $\widetilde{\lambda}_1, \ldots, \widetilde{\lambda}_d$ of $\widetilde{\boldsymbol{A}}$ are the solutions to the secular equation

$$f(\lambda) = 1 + \sum_{i=1}^d \frac{\alpha_i^2}{\lambda_i - \lambda} = 0.$$

2. The eigenvector $\widetilde{\boldsymbol{v}}_1, \ldots, \widetilde{\boldsymbol{v}}_d$ of $\widetilde{\boldsymbol{A}}$ are unit vectors that satisfies

$$\widetilde{\boldsymbol{v}}_i \propto \sum_{j=1}^d \frac{\alpha_i}{\lambda_j - \widetilde{\lambda}_i} \boldsymbol{v}_j.$$

A direct consequence of the rank-one update lemma is listed as follows, where we quantify the growth of the largest eigenvalue of $\boldsymbol{\Lambda}_t$.

Lemma 6 (Growth of the largest eigenvalue)**.** *When $\|\boldsymbol{v}_{t,1} - \widehat{\boldsymbol{\theta}}_t\| \lesssim \beta/\sqrt{\lambda_{t,d}}$, the largest eigenvalue of $\boldsymbol{\Lambda}_t$ evolves according to the update rule*

$$\lambda_{t+1,1} = \lambda_{t,1} + \kappa_{t,1}^2 + O(t^{-1}) = \lambda_{t,1} + 1 - O(\beta^2/\lambda_{t,d}).$$

Multivariate Martingale Lindeberg CLT. A multivariate Lindeberg Central Limit Theorem (CLT) for triangular arrays says that sums of many small, row-wise independent random vectors converge in distribution to a multivariate normal—provided their covariances stabilize and no single term has too much mass in its tails. We present this result as a lemma as follows.

Lemma 7. *Let $\{\boldsymbol{X}_{n,k}\}$ ($n \in \mathbb{N}$, $1 \le k \le m_n$) be an array of $\mathbb{R}^d$-valued random vectors. For each $n \in \mathbb{N}$, let $(\mathcal{F}_{n,k})_{1 \le k \le m_n}$ be a filtration to which $\{\boldsymbol{X}_{n,k}\}_{1 \le k \le m_n}$ is adapted, and suppose the array satisfies the following properties.*

$$\mathbb{E}[\boldsymbol{X}_{n,k}|\mathcal{F}_{n,k-1}] = 0, \quad \boldsymbol{V}_n = \sum_{k=1}^{m_n} \mathrm{Var}(\boldsymbol{X}_{n,k}|\mathcal{F}_{n,k-1}) \to \boldsymbol{\Sigma},$$

where $\boldsymbol{\Sigma}$ is a fixed, positive definite $d \times d$ matrix. Furthermore, the array $\{\boldsymbol{X}_{n,k}\}$ satisfies Lyapunov condition, i.e. there exists $\delta > 0$ such that

$$\sum_{k=1}^{m_n} \mathbb{E}\left[\|\boldsymbol{X}_{n,k}\|^{2+\delta}|\mathcal{F}_{n,k-1}\right] \to 0.$$

Define the row sum $\boldsymbol{S}_n = \sum_{k=1}^{m_n} \boldsymbol{X}_{n,k}$. Then we have

$$\boldsymbol{S}_n \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Sigma}).$$

The multivariate CLT with the Lyapunov condition stated above can be established for example by applying the one-dimensional Lindeberg CLT (Theorem 27.3 in Billingsley (2013)) for one-dimensional projection $\boldsymbol{\theta}^\top \boldsymbol{X}_{n,k}$ ($\boldsymbol{\theta} \in \mathbb{R}^d$) and then using the Cramér–Wold theorem to conclude convergence in distribution of the vector.

B Proof of Theorem 1

We present the proof of Theorem 1, based on the conclusion of Theorem 3. Without loss of generality, throughout the proof, we assume that $\boldsymbol{\theta}^\star = \boldsymbol{e}_1$. For general $\boldsymbol{\theta}^\star$, the result follows from the same analysis.

Step 1: show that $\|\widetilde{\boldsymbol{\Sigma}}_T^{-1}\widetilde{\boldsymbol{\Lambda}}_T - \boldsymbol{I}_{d-1}\|_2 \to 0$. Recall in (30), we decompose $\boldsymbol{\Lambda}_T$ as

$$\boldsymbol{\Lambda}_T = \lambda_{T,1}\boldsymbol{v}_{T,1}\boldsymbol{v}_{T,1}^\top + \sum_{i=2}^d \lambda_{T,i}\boldsymbol{v}_{T,i}\boldsymbol{v}_{T,i}^\top.$$

From Proposition 4, when $\beta \gg d^2(\sigma\sqrt{d + \log\log T} + 1)$, the non-leading eigenvalues satisfy

$$\lambda_{T,i} = (1 + \Delta_{T,i})\sqrt{\frac{2\beta^2 T}{d+1}}, \qquad \forall\, i \ge 2.$$

with the size of $\Delta_{T,i}$ obeying (29), and consequently, the leading eigenvalue $\lambda_{T,1}$ can be characterized as

$$\lambda_{T,1} = \sum_{i=1}^d \lambda_{T,i} - \sum_{i=2}^d \lambda_{T,i} = T + d - \left(d - 1 + \sum_{i=2}^d \Delta_{T,i}\right)\sqrt{\frac{2\beta^2 T}{d+1}}.$$

As a result, we can express $\boldsymbol{\Lambda}_T$ as follows

$$\begin{aligned}
\boldsymbol{\Lambda}_T &= \lambda_{T,1}\boldsymbol{v}_{T,1}\boldsymbol{v}_{T,1}^\top + \sum_{i=2}^d \lambda_{T,i}\boldsymbol{v}_{T,i}\boldsymbol{v}_{T,i}^\top \\
&= \lambda_{T,1}\boldsymbol{v}_{T,1}\boldsymbol{v}_{T,1}^\top + \sqrt{\frac{2\beta^2 T}{d+1}}\sum_{i=2}^d \boldsymbol{v}_{T,i}\boldsymbol{v}_{T,i}^\top + \sqrt{\frac{2\beta^2 T}{d+1}}\sum_{i=2}^d \Delta_{T,i}\boldsymbol{v}_{T,i}\boldsymbol{v}_{T,i}^\top
\end{aligned}$$

$$= \sqrt{\frac{2\beta^2 T}{d+1}} \boldsymbol{I}_d + \left(\lambda_{T,1} - \sqrt{\frac{2\beta^2 T}{d+1}}\right) \boldsymbol{v}_{T,1} \boldsymbol{v}_{T,1}^\top + \sqrt{\frac{2\beta^2 T}{d+1}} \sum_{i=2}^d \Delta_{T,i} \boldsymbol{v}_{T,i} \boldsymbol{v}_{T,i}^\top, \tag{44}$$

Here, as we set

$$\widetilde{\boldsymbol{\Sigma}}_T = \sqrt{\frac{2\beta^2 T}{d+1}} \boldsymbol{I}_{d-1},$$

we can write

$$\widetilde{\boldsymbol{\Lambda}}_T - \widetilde{\boldsymbol{\Sigma}}_T = \left(\lambda_{T,1} - \sqrt{\frac{2\beta^2 T}{d+1}}\right) \cdot \boldsymbol{U}^\top \boldsymbol{v}_{T,1} \boldsymbol{v}_{T,1}^\top \boldsymbol{U} + \sqrt{\frac{2\beta^2 T}{d+1}} \boldsymbol{U}^\top \left(\sum_{i=2}^d \Delta_{T,i} \boldsymbol{v}_{T,i} \boldsymbol{v}_{T,i}^\top\right) \boldsymbol{U}. \tag{45}$$

We proceed to calculate the following quantities

$$\left\|\boldsymbol{U}^\top \boldsymbol{v}_{T,1} \boldsymbol{v}_{T,1}^\top \boldsymbol{U}\right\|_2 = \|\boldsymbol{U}^\top \boldsymbol{v}_{T,1}\|_2^2 \leq \|\boldsymbol{v}_{T,1} - \boldsymbol{\theta}^\star\|_2^2 \lesssim \frac{(\sigma\sqrt{d + \log\log T} + 1)^2}{\lambda_{T,d}}, \tag{46}$$

$$\left\|\sum_{i=2}^d \Delta_{T,i} \boldsymbol{v}_{T,i} \boldsymbol{v}_{T,i}^\top\right\|_2 \lesssim \max_{2 \leq i \leq d} |\Delta_{T,i}| \cdot \left\|\sum_{i=1}^d \boldsymbol{v}_{T,i} \boldsymbol{v}_{T,i}^\top\right\|_2 = \max_{2 \leq i \leq d} |\Delta_{T,i}|, \tag{47}$$

where the first inequality of (46) holds true as

$$\|\boldsymbol{U}^\top \boldsymbol{v}_{T,1}\|_2^2 \leq \left(\|\boldsymbol{U}^\top \boldsymbol{\theta}^\star\|_2 + \|\boldsymbol{U}^\top(\boldsymbol{v}_{T,1} - \boldsymbol{\theta}^\star)\|_2\right)^2 = \|\boldsymbol{U}^\top(\boldsymbol{v}_{T,1} - \boldsymbol{\theta}^\star)\|_2^2 \leq \|\boldsymbol{v}_{T,1} - \boldsymbol{\theta}^\star\|_2^2.$$

Consequently, with (46) and (47), we can upper bound the norm of (45) as

$$\|\widetilde{\boldsymbol{\Lambda}}_T - \widetilde{\boldsymbol{\Sigma}}_T\|_2 \lesssim T \cdot \frac{(\sigma\sqrt{d + \log\log T} + 1)^2}{\lambda_{T,d}} + \sqrt{\frac{2\beta^2 T}{d+1}} \cdot \max_{2 \leq i \leq T} |\Delta_{T,i}|,$$

which leads to the following upper bound when $\beta \gg d^2(\sigma\sqrt{d + \log\log T} + 1)$

$$\begin{aligned}
\left\|\widetilde{\boldsymbol{\Sigma}}_T^{-1} \widetilde{\boldsymbol{\Lambda}}_T - \boldsymbol{I}_{d-1}\right\|_2 &\leq \|\widetilde{\boldsymbol{\Sigma}}_T^{-1}\|_2 \cdot \|\widetilde{\boldsymbol{\Lambda}}_T - \widetilde{\boldsymbol{\Sigma}}_T\|_2 \\
&\lesssim T\sqrt{\frac{d+1}{2\beta^2 T}} \cdot \frac{(\sigma\sqrt{d + \log\log T} + 1)^2}{\lambda_{T,d}} + \max_{2 \leq i \leq d} |\Delta_{T,i}| \\
&\lesssim \frac{d(\sigma\sqrt{d + \log\log T} + 1)^2}{\beta^2} + d\left(\frac{\beta^8}{T\sigma^6}\right)^{\frac{d+1}{d-1}} + \frac{(\sigma\sqrt{d + \log\log T} + 1)^{1/2}}{\sqrt{\beta}} \\
&\lesssim d\left(\frac{\beta^8}{T\sigma^6}\right)^{\frac{d+1}{d-1}} + \frac{(\sigma\sqrt{d + \log\log T} + 1)^{1/2}}{\sqrt{\beta}},
\end{aligned} \tag{48}$$

where the third inequality follows from (29) and the fact that $\lambda_{T,d} \asymp \sqrt{\frac{2\beta^2 T}{d+1}}$, which establishes that

$$\left\|\widetilde{\boldsymbol{\Sigma}}_T^{-1} \widetilde{\boldsymbol{\Lambda}}_T - \boldsymbol{I}_{d-1}\right\|_2 = o(1),$$

whenever $\beta \gg d^2(\sigma\sqrt{d + \log\log T} + 1)$.

Step 2: Asymptotic normality of $\widetilde{\boldsymbol{\Sigma}}_T^{1/2} \boldsymbol{U}^\top(\overline{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star)$. Before diving into the analysis for $\widehat{\boldsymbol{\theta}}_T$, we first deal with the asymptotic of $\overline{\boldsymbol{\theta}}_T$ (defined in (8)), which does not require projection to the unit sphere. To this end, we decompose $\overline{\boldsymbol{\theta}}_T$ as

$$\overline{\boldsymbol{\theta}}_T = \boldsymbol{\Lambda}_T^{-1}\left[(\boldsymbol{\Lambda}_T - \boldsymbol{I}_d)\boldsymbol{\theta}^\star + \boldsymbol{\eta}_T\right] = \boldsymbol{\theta}^\star - \boldsymbol{\Lambda}_T^{-1}\boldsymbol{\theta}^\star + \boldsymbol{\Lambda}_T^{-1}\boldsymbol{\eta}_T, \tag{49}$$

where $\boldsymbol{\eta}_T = \sum_{t=1}^T \boldsymbol{a}_t \epsilon_t$. In terms of this decomposition, we write

$$\widetilde{\boldsymbol{\Sigma}}_T^{1/2} \boldsymbol{U}^\top(\overline{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star) = \widetilde{\boldsymbol{\Sigma}}_T^{1/2} \boldsymbol{U}^\top \boldsymbol{\Lambda}_T^{-1} \boldsymbol{\eta}_T - \widetilde{\boldsymbol{\Sigma}}_T^{1/2} \boldsymbol{U}^\top \boldsymbol{\theta}^\star. \tag{50}$$

We shall deal with these two terms separately. Let us start with the first term since it is more involved. Construct a diagonal matrix

$$\boldsymbol{\Sigma}_T = \sqrt{\frac{2\beta^2 T}{d+1}} \boldsymbol{I}_d. \tag{51}$$

Intuitively, $\boldsymbol{\Sigma}_T$ approximates $\boldsymbol{\Lambda}_T$, except for the first row and column. We decompose the first term as follows

$$\widetilde{\boldsymbol{\Sigma}}_T^{1/2} \boldsymbol{U}^\top \boldsymbol{\Lambda}_T^{-1} \boldsymbol{\eta}_T = \widetilde{\boldsymbol{\Sigma}}_T^{1/2} \boldsymbol{U}^\top \boldsymbol{\Sigma}_T^{-1} \boldsymbol{\eta}_T + \widetilde{\boldsymbol{\Sigma}}_T^{1/2} \boldsymbol{U}^\top (\boldsymbol{\Lambda}_T^{-1} - \boldsymbol{\Sigma}_T^{-1}) \boldsymbol{\eta}_T. \tag{52}$$

We will first derive the asymptotic of the first term above. Here we need to use the triangular array argument, which was stated in Lemma 7. Formally, let $\{\boldsymbol{a}_{t,s}\}$ and $\{\epsilon_{t,s}\}$ ($t \in \mathbb{N}, s \le t$) be arrays of actions and noise when implementing LinUCB. Then we rewrite $\boldsymbol{\Lambda}_T$ and $\boldsymbol{\eta}_T$ in the following way

$$\boldsymbol{\Lambda}_T = \boldsymbol{I}_d + \sum_{s=1}^t \boldsymbol{a}_{T,s} \boldsymbol{a}_{T,s}^\top, \qquad \boldsymbol{\eta}_T = \sum_{s=1}^t \boldsymbol{a}_{T,s} \epsilon_{T,s}. \tag{53}$$

Let $\mathcal{F}_{T,s} = \sigma(\boldsymbol{a}_{T,1}, \epsilon_{T,1}, \ldots, \boldsymbol{a}_{T,s}, \epsilon_{T,s})$ and define

$$\boldsymbol{X}_{T,s} = \widetilde{\boldsymbol{\Sigma}}_T^{1/2} \boldsymbol{U}^\top \boldsymbol{\Sigma}_T^{-1} \boldsymbol{a}_{T,s} \epsilon_{T,s}. \tag{54}$$

Consequently, $\sum_{s=1}^T \boldsymbol{X}_{T,s} = \widetilde{\boldsymbol{\Sigma}}_T^{1/2} \boldsymbol{U}^\top \boldsymbol{\Sigma}_T^{-1} \boldsymbol{\eta}_T$. In addition, $\boldsymbol{X}_{T,s} \in \mathcal{F}_{T,s}$, i.e. $\boldsymbol{X}_{T,s}$ is adapted to the filtration $\mathcal{F}_{T,s}$, and the mean and variance of $\boldsymbol{X}_{T,s}$ conditioned on $\mathcal{F}_{T,s-1}$ are given as

$$\mathbb{E}[\boldsymbol{X}_{T,s}|\mathcal{F}_{T,s-1}] = 0, \quad \text{Var}[\boldsymbol{X}_{T,s}|\mathcal{F}_{T,s-1}] = \sigma^2 \widetilde{\boldsymbol{\Sigma}}_T^{1/2} \boldsymbol{U}^\top \boldsymbol{\Sigma}_T^{-1} \boldsymbol{a}_{T,s} \boldsymbol{a}_{T,s}^\top \boldsymbol{\Sigma}_T^{-1} \boldsymbol{U} \widetilde{\boldsymbol{\Sigma}}_T^{1/2}. \tag{55}$$

As a matter of fact, it holds that

$$\begin{aligned}
\sum_{s=1}^T \text{Var}[\boldsymbol{X}_{T,s}|\mathcal{F}_{T,s-1}] &= \sigma^2 \widetilde{\boldsymbol{\Sigma}}_T^{1/2} \boldsymbol{U}^\top \boldsymbol{\Sigma}_T^{-1} (\boldsymbol{\Lambda}_T - \boldsymbol{I}_d) \boldsymbol{\Sigma}_T^{-1} \boldsymbol{U} \widetilde{\boldsymbol{\Sigma}}_T^{1/2} \\
&= \sigma^2 (\boldsymbol{0}_{d-1}, \widetilde{\boldsymbol{\Sigma}}_T^{-1/2})(\boldsymbol{\Lambda}_T - \boldsymbol{I}_d)(\boldsymbol{0}_{d-1}, \widetilde{\boldsymbol{\Sigma}}_T^{-1/2})^\top \\
&= \sigma^2 \sqrt{\frac{d+1}{2\beta^2 T}} \boldsymbol{U}^\top (\boldsymbol{\Lambda}_T - \boldsymbol{I}_d) \boldsymbol{U}.
\end{aligned} \tag{56}$$

Here, since

$$\sqrt{\frac{d+1}{2\beta^2 T}} \boldsymbol{U}^\top \boldsymbol{\Lambda}_T \boldsymbol{U} = \widetilde{\boldsymbol{\Sigma}}_T^{-1} \widetilde{\boldsymbol{\Lambda}}_T \longrightarrow \boldsymbol{I}_d, \tag{57}$$

which was shown in Step 1, we can conclude

$$\sum_{s=1}^T \text{Var}[\boldsymbol{X}_{T,s}|\mathcal{F}_{T,s-1}] \longrightarrow \sigma^2 \boldsymbol{I}_d. \tag{58}$$

To apply the Martingale Lindeberg CLT (Lemma 7), let us further verify the Lyapunov condition. For any $\delta > 0$, we have

$$\sum_{s=1}^T \mathbb{E}[\|\boldsymbol{X}_{t,s}\|^{2+\delta}|\mathcal{F}_{t,s-1}] = \max_{1 \le s \le t} \mathbb{E}\left[\|\boldsymbol{X}_{t,s}\|^\delta|\mathcal{F}_{t,s-1}\right] \cdot \sum_{s=1}^t \mathbb{E}\left[\|\boldsymbol{X}_{t,s}\|^2|\mathcal{F}_{t,s-1}\right] \le \max_{1 \le s \le t} \mathbb{E}\left[\|\boldsymbol{X}_{t,s}\|^\delta|\mathcal{F}_{t,s-1}\right].$$

We upper bound $\mathbb{E}\left[\|\boldsymbol{X}_{t,s}\|^\delta|\mathcal{F}_{t,s-1}\right]$ by

$$\mathbb{E}\left[\|\boldsymbol{X}_{T,s}\|^\delta|\mathcal{F}_{T,s-1}\right] \lesssim \left\|\widetilde{\boldsymbol{\Sigma}}_T^{1/2} \boldsymbol{U}^\top \boldsymbol{\Sigma}_T^{-1} \boldsymbol{a}_{T,s} \epsilon_{T,s}\right\|^\delta \lesssim \left(\sqrt{\frac{d+1}{2\beta^2 T}}\right)^{\delta/2} \longrightarrow 0,$$

as $T \to \infty$, since $\lambda_{t,d} \to \infty$. Then, by virtue of the Martingale Lindeberg CLT (Lemma 7), we obtain

$$\widetilde{\Sigma}_T^{1/2} U^\top \Sigma_T^{-1} \eta_T \longrightarrow \mathcal{N}(0, \sigma^2 I_d). \tag{59}$$

Let us then consider the term $\widetilde{\Sigma}_T^{1/2} U^\top (\Lambda_T^{-1} - \Sigma_T^{-1}) \eta_T$. In view of the Cauchy-Schwarz inequality, the term of interest can be written as

$$\| U^\top (\Lambda_T^{-1} - \Sigma_T^{-1}) \eta_T \|_2 \leq \left\| U^\top (\Lambda_T^{-1} - \Sigma_T^{-1}) \Lambda_T^{1/2} \right\|_2 \cdot \left\| \Lambda_T^{-1/2} \eta_T \right\|_2. \tag{60}$$

To bound the right hand side of the above inequality, we first note that

$$\mathbb{E}\left[\left\| \Lambda_T^{-1/2} \eta_T \right\|_2^2 \right] = \sigma^2 \operatorname{tr}\left(\Lambda_T^{-1/2} \Big(\sum_{s=1}^{T} a_s a_s^\top \Big) \Lambda_T^{-1/2} \right) \leq \sigma^2 d,$$

which leads to $\| \Lambda_T^{-1/2} \eta_T \|_2 = O_p(1)$. It is therefore only left for us to control the quantity $\| U^\top (\Lambda_T^{-1} - \Sigma_T^{-1}) \Lambda_T^{1/2} \|_2$. Recalling the definition $\Sigma_T^{-1} = \sqrt{\frac{d+1}{2\beta^2 T}} I_d$, we write

$$\left\| U^\top (\Lambda_T^{-1} - \Sigma_T^{-1}) \Lambda_T^{1/2} \right\|_2 = \left\| U^\top \left(\Lambda_T^{-1/2} - \sqrt{\frac{d+1}{2\beta^2 T}} \Lambda_T^{1/2} \right) \right\|_2. \tag{61}$$

To control the right-hand side of the above equality, invoking the decomposition (30) yields that

$$\Lambda_T^{-1/2} - \sqrt{\frac{d+1}{2\beta^2 T}} \Lambda_T^{1/2} = \sum_{i=1}^{d} \left(\frac{1}{\sqrt{\lambda_{T,i}}} - \sqrt{\frac{d+1}{2\beta^2 T}} \sqrt{\lambda_{T,i}} \right) v_{T,i} v_{T,i}^\top. \tag{62}$$

For $i \geq 2$, notice that $\lambda_{T,i} = (1 + o(1))\sqrt{(d+1)/(2\beta^2 T)}$. As a result, we have

$$\frac{1}{\sqrt{\lambda_{T,i}}} - \sqrt{\frac{d+1}{2\beta^2 T}} \sqrt{\lambda_{T,i}} = \frac{1}{\sqrt{\lambda_{T,i}}} \cdot \left(1 - \sqrt{\frac{d+1}{2\beta^2 T}} \lambda_{T,i} \right) = o(1) \cdot \frac{1}{\sqrt{\lambda_{T,i}}}.$$

For $i = 1$, recall

$$\| U^\top v_{T,1} \|_2 \leq \| v_{T,1} - \theta^\star \|_2 \lesssim \frac{\sigma \sqrt{d + \log \log T} + 1}{\sqrt{\lambda_{T,d}}},$$

to arrive at

$$\left\| U^\top \left(\frac{1}{\sqrt{\lambda_{T,1}}} - \sqrt{\frac{d+1}{2\beta^2 T}} \cdot \sqrt{\lambda_{T,1}} \right) v_{T,1} v_{T,1}^\top \right\|_2 \lesssim \frac{\sigma \sqrt{d + \log \log T} + 1}{\sqrt{\lambda_{T,d}}} \cdot \left(\frac{1}{\sqrt{T}} + \sqrt{T} \cdot \frac{\sqrt{d}}{\beta \sqrt{T}} \right)$$

$$\lesssim \frac{\sigma \sqrt{d + \log \log T} + 1}{\sqrt{\lambda_{T,d}}} \cdot \frac{\sqrt{d}}{\beta}, \tag{63}$$

where the last inequality holds as $\beta = O(\operatorname{poly} \log T) \ll T$. Putting pieces together, we obtain the upper bound

$$\left\| U^\top (\Lambda_T^{-1} - \Sigma_T^{-1}) \Lambda_T^{1/2} \right\|_2 \lesssim \frac{1}{\sqrt{\lambda_{T,d}}} \left(d \cdot o(1) + \frac{\sqrt{d}(\sigma \sqrt{d + \log \log T} + 1)}{\beta} \right) = \frac{1}{\sqrt{\lambda_{T,d}}} \cdot o(1), \tag{64}$$

as we set $\beta \gg d^2 (\sigma \sqrt{d + \log \log T} + 1)$. As a result, we conclude that

$$\left\| \widetilde{\Sigma}_T^{1/2} U^\top (\Lambda_T^{-1} - \Sigma_T^{-1}) \eta_T \right\|_2 = \left(\frac{2\beta^2 T}{d+1} \right)^{1/4} \cdot \frac{1}{\sqrt{\lambda_{T,d}}} \cdot o(1) \cdot O_p(1) = o_p(1). \tag{65}$$

In other words, the term $\|\widetilde{\boldsymbol{\Sigma}}_T^{1/2}\boldsymbol{U}^\top(\boldsymbol{\Lambda}_T^{-1} - \boldsymbol{\Sigma}_T^{-1})\boldsymbol{\eta}_T\|_2$ converges to 0 in probability. As a result, one has

$$\widetilde{\boldsymbol{\Sigma}}_T^{1/2}\boldsymbol{U}^\top\boldsymbol{\Lambda}_T^{-1}\boldsymbol{\eta}_T \longrightarrow \mathcal{N}(0, \sigma^2\boldsymbol{I}_d). \tag{66}$$

For the second term, we note that

$$\left\|\widetilde{\boldsymbol{\Sigma}}_T^{1/2}\boldsymbol{U}^\top\boldsymbol{\Lambda}_T^{-1}\boldsymbol{\theta}^\star\right\|_2 \leq \left\|\widetilde{\boldsymbol{\Sigma}}_T^{1/2}\boldsymbol{U}^\top\boldsymbol{\Lambda}_T^{-1}\right\|_2 \leq \left(\frac{2\beta^2 T}{d+1}\right)^{1/4} \cdot \frac{1}{\lambda_{T,d}} \longrightarrow 0,$$

as $T \to \infty$. Combining this with (50) and (66), we arrive at

$$\widetilde{\boldsymbol{\Sigma}}_T^{1/2}\boldsymbol{U}^\top(\overline{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star) \longrightarrow \mathcal{N}(0, \sigma^2\boldsymbol{I}_d). \tag{67}$$

Step 3: Asymptotic normality of $\widetilde{\boldsymbol{\Sigma}}_T^{1/2}\boldsymbol{U}^\top(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star)$. In view of the decomposition in (49), we bound the norm of $\overline{\boldsymbol{\theta}}_T$ by triangle's inequality as

$$\left|\|\overline{\boldsymbol{\theta}}_T\|_2 - 1\right| \leq \left\|\boldsymbol{\Lambda}_T^{-1}\boldsymbol{\theta}^\star\right\|_2 + \left\|\boldsymbol{\Lambda}_T^{-1}\boldsymbol{\eta}_T\right\|_2 \leq \frac{1}{\lambda_{T,d}} + \frac{1}{\sqrt{\lambda_{T,d}}}\left\|\boldsymbol{\Lambda}_T^{-1/2}\boldsymbol{\eta}_T\right\|_2 = O_p\left(\frac{1}{\sqrt{\lambda_{T,d}}}\right). \tag{68}$$

As a result, for the projection $\widehat{\boldsymbol{\theta}}_T = \overline{\boldsymbol{\theta}}_T/\|\overline{\boldsymbol{\theta}}_T\|_2$, it is easily seen that

$$\|\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star\| \leq \|\overline{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star\| = O_p\left(\frac{1}{\sqrt{\lambda_{T,d}}}\right). \tag{69}$$

Furthermore, as $\widehat{\boldsymbol{\theta}}_T$ being the projection of $\overline{\boldsymbol{\theta}}_T$, we have

$$(\boldsymbol{\theta}^\star)^\top(\overline{\boldsymbol{\theta}}_T - \widehat{\boldsymbol{\theta}}_T) = (\boldsymbol{\theta}^\star)^\top\widehat{\boldsymbol{\theta}}_T \cdot \|\overline{\boldsymbol{\theta}}_T - \widehat{\boldsymbol{\theta}}_T\|_2.$$

Observing the relation

$$\left\|(\boldsymbol{\theta}^\star)^\top(\overline{\boldsymbol{\theta}}_T - \widehat{\boldsymbol{\theta}}_T)\right\|_2^2 + \left\|\boldsymbol{U}^\top\left(\overline{\boldsymbol{\theta}}_T - \widehat{\boldsymbol{\theta}}_T\right)\right\|_2^2 = \|\overline{\boldsymbol{\theta}}_T - \widehat{\boldsymbol{\theta}}_T\|_2^2,$$

leads to

$$\left\|\boldsymbol{U}^\top\left(\overline{\boldsymbol{\theta}}_T - \widehat{\boldsymbol{\theta}}_T\right)\right\|_2 = \sqrt{1 - \left[(\boldsymbol{\theta}^\star)^\top\widehat{\boldsymbol{\theta}}_T\right]^2} \cdot \|\overline{\boldsymbol{\theta}}_T - \widehat{\boldsymbol{\theta}}_T\|_2 \lesssim \|\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star\| \cdot \|\overline{\boldsymbol{\theta}}_T - \widehat{\boldsymbol{\theta}}_T\|_2$$

$$= O_p\left(\frac{1}{\lambda_{T,d}}\right), \tag{70}$$

where the first inequality holds as

$$\left\|\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star\right\|_2^2 = 2\left(1 - \langle\widehat{\boldsymbol{\theta}}_T, \boldsymbol{\theta}^\star\rangle\right) \geq \left(1 - \langle\widehat{\boldsymbol{\theta}}_T, \boldsymbol{\theta}^\star\rangle\right)\left(1 + \langle\widehat{\boldsymbol{\theta}}_T, \boldsymbol{\theta}^\star\rangle\right) = 1 - \langle\widehat{\boldsymbol{\theta}}_T, \boldsymbol{\theta}^\star\rangle^2.$$

Therefore, we obtain the following result

$$\left(\frac{2\beta^2 T}{d+1}\right)^{1/4}\boldsymbol{U}^\top\left(\overline{\boldsymbol{\theta}}_T - \widehat{\boldsymbol{\theta}}_T\right) = O_p\left(\frac{1}{\sqrt{\lambda_{T,d}}}\right). \tag{71}$$

Since $\lambda_{T,d} \to \infty$, combining this with the asymptotic result in Step 2, we conclude that

$$\left(\frac{2\beta^2 T}{d+1}\right)^{1/4}\boldsymbol{U}^\top\left(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}^\star\right) \to \mathcal{N}(0, \boldsymbol{I}_{d-1}).$$

C Proof of Theorem 2

We present the full proof of Theorem 2 in this section. Although a similar result was established in Lemma 5.1 of Khamaru and Zhang (2024) based on a previous result (see this blog), the proof in our setting is much more involved. This increased complexity arises from two key challenges: first, the noise term $\boldsymbol{\eta}_t = \sum_{s=1}^{t} \boldsymbol{a}_s \epsilon_s$ is multi-dimensional, and second, the sample covariance matrix $\boldsymbol{\Lambda}_t$ evolves in a non-stationary manner over time. To proceed, let us denote

$$\widetilde{\boldsymbol{\theta}}_t = \mathbb{E}\left[\left(\sum_{s=1}^{t-1} \boldsymbol{a}_s \boldsymbol{a}_s^T + \boldsymbol{I}_d\right)^{-1} \left(\sum_{s=1}^{t-1} \boldsymbol{a}_s y_s\right)\right] = \boldsymbol{\Lambda}_{t-1}^{-1}(\boldsymbol{\Lambda}_{t-1} - \boldsymbol{I}_d)\boldsymbol{\theta}^\star, \tag{72}$$

be the expectation of ridge regression regression estimator $\overline{\boldsymbol{\theta}}_t$, and set $\boldsymbol{\eta}_t = \sum_{s=1}^{t} \boldsymbol{a}_s \epsilon_s$ be the corresponding noise part. Then we can rewrite $\overline{\boldsymbol{\theta}}_t$ as

$$\overline{\boldsymbol{\theta}}_t = \boldsymbol{\Lambda}_{t-1}^{-1}\left(\sum_{s=1}^{t-1} \boldsymbol{a}_s y_s\right) = \widetilde{\boldsymbol{\theta}}_t + \boldsymbol{\Lambda}_{t-1}^{-1}\boldsymbol{\eta}_{t-1}. \tag{73}$$

Note that $\widetilde{\boldsymbol{\theta}}_t$ is the estimator of ridge regression when no noise is present. We can easily bound the difference between $\widetilde{\boldsymbol{\theta}}_t$ and $\boldsymbol{\theta}^\star$ as follows:

$$\widetilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^\star = \boldsymbol{\Lambda}_{t-1}^{-1}(\boldsymbol{\Lambda}_{t-1} - \boldsymbol{I}_d)\boldsymbol{\theta}^\star - \boldsymbol{\theta}^\star = -\boldsymbol{\Lambda}_{t-1}^{-1}\boldsymbol{\theta}^\star, \tag{74}$$

then one has

$$\left\|\boldsymbol{\Lambda}_t^{1/2}(\widetilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^\star)\right\|_2 \leq \frac{1}{\sqrt{\lambda_{t,d}}}. \tag{75}$$

It remains for one to establish

$$\max_{1 \leq t \leq T} \left\|\boldsymbol{\Lambda}_t^{-1/2}\boldsymbol{\eta}_t\right\|_2 \lesssim \sigma\sqrt{d + \log\log T}.$$

Towards this, we begin by observing the following identity:

$$\frac{1}{2\sigma^2}\left\|\boldsymbol{\Lambda}_t^{-1/2}\boldsymbol{\eta}_t\right\|_2^2 = \frac{1}{2\sigma^2}\boldsymbol{\eta}_t^T\boldsymbol{\Lambda}_t^{-1}\boldsymbol{\eta}_t = \max_{\boldsymbol{\lambda} \in \mathbb{R}^d}\left\{\boldsymbol{\lambda}^T\boldsymbol{\eta}_t - \frac{\sigma^2}{2}\boldsymbol{\lambda}^T\boldsymbol{\Lambda}_t\boldsymbol{\lambda}\right\}. \tag{76}$$

Consequently, in order to control the left-hand side of (76)—which plays a key role in our proof—it suffices to control the right-hand side, that is, the supremum over $\boldsymbol{\lambda} \in \mathbb{R}^d$ of the random process $\boldsymbol{\lambda}^T\boldsymbol{\eta}_t - \frac{\sigma^2}{2}\boldsymbol{\lambda}^T\boldsymbol{\Lambda}_t\boldsymbol{\lambda}$.

To facilitate this, we introduce the following exponential process indexed by $\boldsymbol{\lambda}$ and adapted to the filtration up to time t:

$$M_t(\boldsymbol{\lambda}) = \exp\left(\boldsymbol{\lambda}^T\boldsymbol{\eta}_t - \frac{\sigma^2}{2}\boldsymbol{\lambda}^T\boldsymbol{\Lambda}_t\boldsymbol{\lambda}\right). \tag{77}$$

Here, we claim that

$$M_t(\boldsymbol{\lambda}) \text{ is a supermartingale for any } \boldsymbol{\lambda} \in \mathbb{R}^d, \tag{78}$$

which is proved at the end of this section. To extend this to a uniform bound over $\boldsymbol{\lambda}$, we follow a strategy similar to that in Khamaru and Zhang (2024) and consider a weighted aggregation of the processes $M_t(\boldsymbol{\lambda})$. Specifically, we define:

$$Z_t = \int \gamma(\boldsymbol{\lambda}) M_t(\boldsymbol{\lambda}) \, d\boldsymbol{\lambda}, \tag{79}$$

where $\gamma(\boldsymbol{\lambda})$ is a prior density (or mass function) over $\mathbb{R}^d$ satisfying $\int \gamma(\boldsymbol{\lambda}) \, d\boldsymbol{\lambda} = 1$. Since $M_t(\boldsymbol{\lambda})$ is a non-negative supermartingale and $\gamma(\boldsymbol{\lambda})$ integrates to one, standard results ensure that Z_t is also a super-martingale.

For analytical convenience, we may take $\gamma(\boldsymbol{\lambda})$ to be a discrete prior supported on a countable set $\{\boldsymbol{\lambda}_i\}_{i=1}^{\infty}$ with weights $\{\gamma_i\}_{i=1}^{\infty}$, where $\sum_{i=1}^{\infty} \gamma_i = 1$. In this case, Z_t admits the form:

$$Z_t = \sum_{i=1}^{\infty} \gamma_i M_t(\boldsymbol{\lambda}_i), \tag{80}$$

which is again a supermartingale as a convex combination of supermartingales. This construction allows us to control the supremum over $\boldsymbol{\lambda} \in \mathbb{R}^d$ via a union bound or concentration argument over the discrete support, thereby paving the way for a high-probability bound on (76). As a matter of fact, by uniform concentration of supermartingale, one may show that

$$P\left(\exists t : Z_t \geq \frac{1}{\delta}\right) \leq \delta. \tag{81}$$

Consider the event $E_{i,t} = \{\gamma_i M_t(\boldsymbol{\lambda}_i) \geq \delta^{-1}\}$. Observe that $E_{i,t} \subset \{Z_t \geq \delta^{-1}\}$, since Z_t is a weighted sum over the $M_t(\boldsymbol{\lambda}_i)$. Taking a union over all indices i and all time steps t, we define the event

$$E := \bigcup_{i=1}^{\infty} \bigcup_{t=1}^{T} E_{i,t} \subset \left\{\exists\, t \in [T] : Z_t \geq \frac{1}{\delta}\right\}, \tag{82}$$

which, by the concentration inequality in (81), implies that $\mathbb{P}(E) \leq \delta$. Moreover, each event $E_{i,t}$ can be equivalently rewritten in terms of an inequality involving $\boldsymbol{\eta}_t$ and $\boldsymbol{\Lambda}_t$:

$$E_{i,t} = \left\{\gamma_i \exp\left(\boldsymbol{\lambda}_i^{\top} \boldsymbol{\eta}_t - \frac{\sigma^2}{2}\boldsymbol{\lambda}_i^{\top} \boldsymbol{\Lambda}_t \boldsymbol{\lambda}_i\right) \geq \frac{1}{\delta}\right\} = \left\{\boldsymbol{\lambda}_i^{\top} \boldsymbol{\eta}_t - \frac{\sigma^2}{2}\boldsymbol{\lambda}_i^{\top} \boldsymbol{\Lambda}_t \boldsymbol{\lambda}_i \geq \log\left(\frac{1}{\gamma_i \delta}\right)\right\}. \tag{83}$$

Conditioned on the complement event E^c, which occurs with probability at least $1-\delta$, none of the events $E_{i,t}$ hold for any i or t. This allows us to uniformly control the values of the process in (77) across the net points $\{\boldsymbol{\lambda}_i\}$.

To translate this control into a bound on (76), we construct a weighted net, denoted by $\mathcal{N} = \{\boldsymbol{\lambda}_i\}$ with associated weights $\{\gamma_i\}$, satisfying the following approximation guarantee: with probability $1-\delta$, for any $(\boldsymbol{\eta}_t, \boldsymbol{\Lambda}_t)$ such that $\|\boldsymbol{\Lambda}_t^{-1/2}\boldsymbol{\eta}_t\|_2 \geq \sigma^2$, there exists an index i_t such that

$$\max_{\boldsymbol{\lambda} \in \mathbb{R}^d}\left\{\boldsymbol{\lambda}^T \boldsymbol{\eta}_t - \frac{B^2}{2}\boldsymbol{\lambda}^T \boldsymbol{\Lambda}_t \boldsymbol{\lambda}\right\} \leq \frac{16}{7}\left(\boldsymbol{\lambda}_{i_t}^T \boldsymbol{\eta}_t - \frac{\sigma^2}{2}\boldsymbol{\lambda}_{i_t}^T \boldsymbol{\Lambda}_t \boldsymbol{\lambda}_{i_t}\right), \quad \text{with} \quad \gamma_{i_t} \gtrsim (13^d \cdot \mathrm{poly}(d) \cdot \mathrm{poly}(\log T))^{-1}. \tag{84}$$

As a result, we conclude that with probability at least $1-\delta$, for every $t \in [T]$, the following bound holds:

$$\|\boldsymbol{\Lambda}_t^{-1/2}\boldsymbol{\eta}_t\|_2 \lesssim \sqrt{2}\sigma \cdot \max\left(\sigma, \sqrt{\log\left(\frac{1}{\gamma_{i_t}\delta}\right)}\right) \lesssim \sigma\sqrt{d + \log\left(\frac{\log T}{\delta}\right)}. \tag{85}$$

Setting $\delta = (\log T)^{-1}$ completes the proof of the desired result.

To complete the proof of Theorem 2, it remains to construct a net that satisfies the covering condition in (84). As previously discussed, the primary difficulty lies in the fact that the pair $(\boldsymbol{\Lambda}_t, \boldsymbol{\eta}_t)$ evolves with time and may exhibit instability, making it challenging to construct a uniform net over all t. To make progress, we first consider a simplified setting where t is fixed.

Building a net for a single pair $(\boldsymbol{\Lambda}_t, \boldsymbol{\eta}_t)$. We begin by constructing a covering net for a single instance of the pair $(\boldsymbol{\Lambda}_t, \boldsymbol{\eta}_t)$, which serves as a foundational step toward addressing the more general case where these quantities vary with time. In this simplified setting, we do not assign weights to the elements of the net $\mathcal{N}_t$; instead, our goal is to control the size of $\mathcal{N}_t$ and ensure that it remains as small as possible while still providing sufficient coverage, i.e. with high probability there exists index i such that

$$\max_{\boldsymbol{\lambda} \in \mathbb{R}^d}\left\{\boldsymbol{\lambda}^T \boldsymbol{\eta}_t - \frac{\sigma^2}{2}\boldsymbol{\lambda}^T \boldsymbol{\Lambda}_t \boldsymbol{\lambda}\right\} \leq \frac{9}{5}\left(\boldsymbol{\lambda}_i^T \boldsymbol{\eta}_t - \frac{\sigma^2}{2}\boldsymbol{\lambda}_i^T \boldsymbol{\Lambda}_t \boldsymbol{\lambda}_i\right), \quad \text{if } \|\boldsymbol{\Lambda}_t^{-1/2}\boldsymbol{\eta}_t\|_2 \geq \sigma^2. \tag{86}$$

29

To help us construct the grid, we claim that for each $(\mathbf{\Lambda}_t, \mathbf{\eta}_t)$, the points that satisfy the above condition form a ball in $\mathbb{R}^d$. We characterize that in the following lemma, which is characterized in the following lemma.

Lemma 8. *Let $f(\mathbf{\lambda}) = \mathbf{\lambda}^T \mathbf{\eta} - (\sigma^2/2)\mathbf{\lambda}^T \mathbf{\Lambda} \mathbf{\lambda}$. For any $\kappa \in (0,1)$, if we define the set $\mathcal{C}$ to be $\mathcal{C} = \{\mathbf{\lambda}_0 : f(\mathbf{\lambda}_0) \geq (1 - \kappa^2) \max_{\mathbf{\lambda} \in \mathbb{R}^d} f(\mathbf{\lambda})\}$. Then $\mathcal{C}$ can also be characterized as*

$$\mathcal{C} = \left\{ \mathbf{\lambda}_0 : \left\| \mathbf{\Lambda}^{1/2} \mathbf{\lambda}_0 - \frac{1}{\sigma^2} \mathbf{\Lambda}^{-1/2} \mathbf{\eta} \right\|_2 \leq \frac{\kappa}{\sigma^2} \left\| \mathbf{\Lambda}^{-1/2} \mathbf{\eta} \right\|_2 \right\}.$$

By Lemma 8, an equivalent condition for approximating the maximizer in our variational bound is the existence of a point $\mathbf{\lambda}_i \in \mathcal{N}_t$ such that

$$\left\| \mathbf{\Lambda}_t^{1/2} \mathbf{\lambda}_i - \frac{1}{B^2} \mathbf{\Lambda}_t^{-1/2} \mathbf{\eta}_t \right\|_2 \leq \frac{2}{3\sigma^2} \left\| \mathbf{\Lambda}_t^{-1/2} \mathbf{\eta}_t \right\|_2, \quad \text{whenever } \|\mathbf{\Lambda}_t^{-1/2} \mathbf{\eta}_t\|_2 \geq \sigma^2. \tag{87}$$

This means that at least one point $\mathbf{\lambda}_i \in \mathcal{N}_t$ lies within a Euclidean ball centered at $\frac{1}{B^2} \mathbf{\Lambda}_t^{-1/2} \mathbf{\eta}_t$ with radius $\frac{1}{2B^2} \|\mathbf{\Lambda}_t^{-1/2} \mathbf{\eta}_t\|_2$. To analyze this further, we state the following lemma that gives us an upper bound of $\|\mathbf{\Lambda}_t^{-1/2} \mathbf{\eta}_t\|_2$.

Lemma 9 (Adapted Version of Theorem 6.3.2, Vershynin (2018))**.** *Let $\mathbf{A}$ be an $n \times d$ matrix and let $\mathbf{X} = (X_1, \ldots, X_n)$ be a random vector with independent mean 0, σ^2-variance and σ sub-gaussian coordinates. Then with probability $1 - \delta$,*

$$\|\mathbf{A}\mathbf{X}\|_2 \lesssim \sigma^2 \left(\|\mathbf{A}\|_F + \sqrt{\log\left(\frac{1}{\delta}\right)} \cdot \|\mathbf{A}\|_{\mathrm{op}} \right).$$

To derive the upper bound of $\|\mathbf{\Lambda}_t^{-1/2} \mathbf{\eta}_t\|_2$ from Lemma 9, define the matrix $\mathbf{A}_t = \mathbf{\Lambda}_t^{-1/2}(\mathbf{a}_1, \ldots, \mathbf{a}_t) \in \mathbb{R}^{d \times t}$ and let $\mathbf{X}_t = (\epsilon_1, \ldots, \epsilon_t)^\top \in \mathbb{R}^t$ denote the vector of noise terms. Noting that $\|\mathbf{A}_t\|_{\mathrm{F}} = \sqrt{d}$ and $\|\mathbf{A}_t\|_{\mathrm{op}} = 1$, we apply Lemma 9 to obtain

$$\left\| \mathbf{\Lambda}_t^{-1/2} \mathbf{\eta}_t \right\|_2 = \|\mathbf{A}_t \mathbf{X}_t\|_2 \lesssim \sigma^2 \sqrt{d + \log\left(\frac{1}{\delta}\right)}, \tag{88}$$

with probability at least $1 - \delta$.

Since this is the only information we have about $\mathbf{\Lambda}_t^{-1/2} \mathbf{\eta}_t$, we shall use this upper bound to guide the construction of the covering net. In particular, the net $\mathcal{N}_t$ must satisfy the following condition: for any vector $\mathbf{x} \in \mathbb{R}^d$ such that $\sigma^2 \leq \|\mathbf{x}\|_2 \lesssim \sigma^2 \sqrt{d + \log(\delta^{-1})}$, there exists a point $\mathbf{\lambda} \in \mathcal{N}_t$ such that

$$\left\| \mathbf{\Lambda}_t^{1/2} \mathbf{\lambda} - \frac{1}{B^2} \mathbf{x} \right\|_2 \leq \frac{2}{3\sigma^2} \|\mathbf{x}\|_2. \tag{89}$$

To construct the net $\mathcal{N}_t$ for a fixed time step t, we proceed in three stages:

- **Step 1: Construct a base ϵ-net on the unit sphere.** We begin by constructing an ϵ-net on the unit Euclidean sphere $\mathcal{S}^{d-1}$ in $\mathbb{R}^d$, where we set $\epsilon = 1/6$. Let $\mathcal{N}_0$ denote this covering. By a standard volume argument (see, e.g., Vershynin (2018)), such a net exists with cardinality bounded by

$$|\mathcal{N}_0| \leq \left(1 + \frac{2}{\epsilon}\right)^d \leq 13^d.$$

  This ensures that every point on the unit sphere lies within Euclidean distance ϵ of some point in $\mathcal{N}_0$.

- **Step 2: Construct nets on expanding spheres.** To cover the full range of norms that may arise in the transformed space (i.e., after rescaling by $\mathbf{\Lambda}_t^{-1/2}$ and B^{-2}), we scale $\mathcal{N}_0$ to form nets on concentric spheres of increasing radii. Specifically, for $j = 0, 1, \ldots, k$ we construct $(3/2)^j \epsilon$ covering on sphere

with radii $(3/2)^j$, where k is chosen to ensure that the largest radius exceeds the typical scale of the transformed vector. In particular, we set

$$k = O\left(\log(d + \log(\delta^{-1}))\right),$$

so that the maximum radius $(3/2)^k$ covers the high-probability upper bound of $\sigma^{-2}\|\mathbf{\Lambda}_t^{-1/2}\boldsymbol{\eta}_t\|_2$.

- **Step 3: Construct the final net in the transformed space.** For each j, we map the scaled points from the unit sphere net through the transformation $\boldsymbol{\lambda} = (3/2)^j \cdot \mathbf{\Lambda}_t^{-1/2}\boldsymbol{v}$, where $\boldsymbol{v} \in \mathcal{N}_0$. This results in a net that discretizes the ellipsoidal region defined by the inverse covariance geometry of $\mathbf{\Lambda}_t$. The final net for time t is thus given by

$$\mathcal{N}_t = \left\{\boldsymbol{\lambda} = 2^j \cdot \mathbf{\Lambda}_t^{-1/2}\boldsymbol{v} : \boldsymbol{v} \in \mathcal{N}_0, \ j = 0, 1, \ldots, k\right\}, \tag{90}$$

to account for the full effective range of relevant $\boldsymbol{x}$ vectors satisfying $\sigma^2 \leq \|\boldsymbol{x}\|_2 \lesssim \sigma^2\sqrt{d + \log(\delta^{-1})}$.

With this construction, the total size of each individual $\mathcal{N}_t$ satisfies

$$N_\delta = k|\mathcal{N}_0| \lesssim 13^d \log(d + \log(\delta^{-1})). \tag{91}$$

We now show that $\mathcal{N}$ satisfies the approximate property stated in (89), which concludes our construction of $\mathcal{N}$ in this case.

Let $j = \left\lfloor \log_2\left(\sigma^{-2}\|\boldsymbol{x}\|_2\right)\right\rfloor$ be the largest index such that $2^j \leq \sigma^{-2}\|\boldsymbol{x}\|_2$. By construction, we include a $2^j\epsilon$-net on the sphere $2^j\mathcal{S}^{d-1}$ in $\mathcal{N}$, where $\epsilon = 1/6$. We first observe that

$$d\left(\sigma^{-2}\boldsymbol{x}, 2^j\mathcal{S}^{d-1}\right) \leq \frac{1}{3\sigma^2}\|\boldsymbol{x}\|_2, \tag{92}$$

where $d(\cdot, \cdot)$ denotes the Euclidean distance in $\mathbb{R}^d$. Let

$$\boldsymbol{z} = \arg\min_{\boldsymbol{y} \in 2^j\mathcal{S}^{d-1}} \left\|\sigma^{-2}\boldsymbol{x} - \boldsymbol{y}\right\|_2.$$

Then $d(\sigma^{-2}\boldsymbol{x}, 2^j\mathcal{S}^{d-1}) = \|\sigma^{-2}\boldsymbol{x} - \boldsymbol{z}\|_2$. Since $\mathcal{N}$ contains a $2^j\epsilon$-net on $2^j\mathcal{S}^{d-1}$, there exists $\boldsymbol{\lambda} \in \mathcal{N}$ such that

$$\|\boldsymbol{z} - \boldsymbol{\lambda}\|_2 \leq \frac{2^j}{6} \leq \frac{1}{6\sigma^2}\|\boldsymbol{x}\|_2. \tag{93}$$

Combining (92) and (93) using the triangle inequality, we obtain

$$\left\|\sigma^{-2}\boldsymbol{x} - \boldsymbol{\lambda}\right\|_2 \leq \left\|\sigma^{-2}\boldsymbol{x} - \boldsymbol{z}\right\|_2 + \|\boldsymbol{z} - \boldsymbol{\lambda}\|_2 \leq \frac{1}{3\sigma^2}\|\boldsymbol{x}\|_2 + \frac{1}{6\sigma^2}\|\boldsymbol{x}\|_2 = \frac{1}{2\sigma^2}\|\boldsymbol{x}\|_2. \tag{94}$$

This proves that $\mathcal{N}$ satisfies the desired approximation.

Building a net with uniform approximation on all $(\mathbf{\Lambda}_t, \boldsymbol{\eta}_t)$. We now turn to the more general case, where the goal is to construct a net that uniformly approximates all pairs $(\mathbf{\Lambda}_t, \boldsymbol{\eta}_t)$ for every $t \in [T]$. A straightforward extension of the previous construction would be to build a separate net $\mathcal{N}_t$ for each t. However, this naive approach yields a total of at least $O(T)$ points in the combined net, implying that some individual points would be assigned weights of order $O(T^{-1})$. This clearly contradicts the desired guarantee in (84). Consequently, we require an alternative construction of a global net $\mathcal{N}$ whose size scales only as $O(\log T)$, while still ensuring uniform approximation across all time indices $t \in [T]$.

Our construction is inspired by rare-switching techniques commonly used in online reinforcement learning. However, in contrast to (He et al., 2023; Tan et al., 2025), where rare-switching is applied to policy updates to control the growth of the function class, we adopt a different perspective. Specifically, we apply the rare-switching principle to select a small number—only $O(\log T)$—of representative time indices $\{t_i\}$, and construct nets $\mathcal{N}_{t_i}$ based solely on the pairs $(\mathbf{\Lambda}_{t_i}, \boldsymbol{\eta}_{t_i})$.

The key insight is that even though we are not covering every $(\mathbf{\Lambda}_t, \boldsymbol{\eta}_t)$ individually, these $O(\log T)$ representative nets are sufficient to uniformly approximate all $(\mathbf{\Lambda}_t, \boldsymbol{\eta}_t)$ across $t \in [T]$. In other words, for any t, there exists some t_i such that the corresponding net $\mathcal{N}_{t_i}$ provides a good approximation for $(\mathbf{\Lambda}_t, \boldsymbol{\eta}_t)$. This significantly reduces the size of the overall net $\mathcal{N}$ while preserving the desired approximation guarantees.

We now describe the construction procedure for the weighted net $\mathcal{N}$, which leverages this rare-switching idea to achieve efficient coverage over all $(\mathbf{\Lambda}_t, \boldsymbol{\eta}_t)$ with only logarithmically many representative components. The construction proceeds as follows:

- **Step 1: Selection of representative time indices.** Initialize $s_1 = 1$, and recursively define the sequence $\{s_2, s_3, \ldots, s_j\}$ according to the rule:

$$s_{i+1} = \min \left\{ t \leq T : \det(\mathbf{\Lambda}_t) > 2 \cdot \det(\mathbf{\Lambda}_{s_i}) \right\}. \tag{95}$$

This procedure continues until no further such t exists. We claim that the following properties hold true for the switching times defined in (95).

$$j = O(d \log T); \quad s_{i+1} - s_i \leq 2^{i-1} d, \ \forall\, 1 \leq i \leq j. \tag{96}$$

The intuition is that each $\mathbf{\Lambda}_{s_i}$ represents a "scale" of covariance growth, and doubling the determinant indicates a substantial geometric change in the feature space. With this setup, we ensure that $\mathbf{\Lambda}_t$ ($s_i \leq t < s_{i+1}$) does not change much from $\mathbf{\Lambda}_{s_i}$.

- **Step 2: Construction of local nets.** For each selected time index s_i, construct a net $\mathcal{N}_{s_i}$ following the procedure described in (90), with $\delta_i = \frac{1}{2^{i-1} d^2 \log^2 T}$. Set $N_i = |\mathcal{N}_{s_i}|$.

- **Step 3: Aggregation into a weighted net.** Let $\mathcal{N}$ be the weighted union of all constructed nets $\{\mathcal{N}_{s_i}\}_{i=1}^{j}$. Assign each net $\mathcal{N}_{s_i}$ a total weight of $\gamma_i = \frac{1}{i(i+1)}$, distributed uniformly across its N points. The remaining probability mass is assigned to the zero vector $\mathbf{0}$ with weight $\gamma_0 = \frac{1}{j+1}$. This ensures the total weight sums to one. The resulting weighted net can be written explicitly as:

$$\mathcal{N} = \left\{ (\overline{\gamma}_i, \boldsymbol{\lambda}_{i,m}) : 1 \leq i \leq j,\ 1 \leq m \leq N_i,\ \overline{\gamma}_i = \frac{1}{i(i+1)N_i},\ \boldsymbol{\lambda}_{i,m} \in \mathcal{N}_{s_i} \right\} \cup \left\{ (\gamma_0, \mathbf{0}) \right\}. \tag{97}$$

We will then show below that the net $\mathcal{N}$ satisfies property (84) for any $(\eta_t, \mathbf{\Lambda}_t)$. By Lemma 8, we only need to show that for any t that satisfies $s_i \leq t < s_{i+1}$, there exists $\boldsymbol{\lambda} \in \mathcal{N}_{s_i}$ such that

$$\left\| \mathbf{\Lambda}_t^{1/2} \boldsymbol{\lambda} - \frac{1}{\sigma^2} \mathbf{\Lambda}_t^{-1/2} \boldsymbol{\eta}_t \right\|_2 \leq \frac{3}{4\sigma^2} \left\| \mathbf{\Lambda}_t^{-1/2} \boldsymbol{\eta}_t \right\|_2, \quad \text{whenever } \|\mathbf{\Lambda}_t^{-1/2} \boldsymbol{\eta}_t\|_2 \geq \sigma^2. \tag{98}$$

As we note that

$$1 \leq \lambda_{\min}(\mathbf{\Lambda}_t^{1/2} \mathbf{\Lambda}_{s_i}^{-1/2}) \leq \lambda_{\max}(\mathbf{\Lambda}_t^{1/2} \mathbf{\Lambda}_{s_i}^{-1/2}) \leq \sqrt{2}, \quad \frac{1}{\sqrt{2}} \leq \lambda_{\min}(\mathbf{\Lambda}_{s_i}^{1/2} \mathbf{\Lambda}_t^{-1/2}) \leq \lambda_{\max}(\mathbf{\Lambda}_{s_i}^{1/2} \mathbf{\Lambda}_t^{-1/2}) \leq 1,$$

conditioned on $\sqrt{2} \leq \sigma^{-2} \|\mathbf{\Lambda}_t^{-1/2} \boldsymbol{\eta}_t\|_2 \leq R$, it holds that

$$\begin{aligned}
\left\| \mathbf{\Lambda}_t^{1/2} \boldsymbol{\lambda} - \frac{1}{B^2} \mathbf{\Lambda}_t^{-1/2} \boldsymbol{\eta}_t \right\|_2 &= \left\| (\mathbf{\Lambda}_t^{1/2} \mathbf{\Lambda}_{s_i}^{-1/2}) \mathbf{\Lambda}_{s_i}^{1/2} \boldsymbol{\lambda} - \frac{1}{\sigma^2} (\mathbf{\Lambda}_t^{1/2} \mathbf{\Lambda}_{s_i}^{-1/2})(\mathbf{\Lambda}_{s_i}^{1/2} \mathbf{\Lambda}_t^{-1/2}) \mathbf{\Lambda}_t^{1/2} \boldsymbol{\eta}_t \right\|_2 \\
&\leq \sqrt{2} \left\| \mathbf{\Lambda}_{s_i}^{1/2} \boldsymbol{\lambda} - \frac{1}{\sigma^2} (\mathbf{\Lambda}_{s_i}^{1/2} \mathbf{\Lambda}_t^{-1/2}) \mathbf{\Lambda}_t^{-1/2} \boldsymbol{\eta}_t \right\|_2 \\
&\leq \sqrt{2} \max_{1 \leq \|\boldsymbol{x}\|_2 \leq R} \left\| \mathbf{\Lambda}_{s_i}^{1/2} \boldsymbol{\lambda} - \boldsymbol{x} \right\|_2.
\end{aligned} \tag{99}$$

Next, we are going to give a uniform upper bound for $\max_{s_i \leq t < s_{i+1}} \sigma^{-2} \|\mathbf{\Lambda}_t^{-1/2} \boldsymbol{\eta}_t\|_2$. We utilize the following lemma to show this result

From Lemma 9, we know that with probability $1 - \frac{1}{2^{i-1}d^2\log^2 T}$, we have

$$\|\boldsymbol{\Lambda}_t^{-1/2}\boldsymbol{\eta}_t\|_2 \lesssim \sigma^2\sqrt{d+i+\log\log T} \lesssim \sigma^2\sqrt{d\log T}. \tag{100}$$

From Claim (96) that $s_{i+1} - s_i \leq 2^{i-1}d$, we obtain that with probability $1 - \frac{1}{d^2\log^2 T}$, one has

$$\max_{s_i \leq t < s_{i+1}} \|\boldsymbol{\Lambda}_t^{-1/2}\boldsymbol{\eta}_t\|_2 \lesssim \sigma^2\sqrt{d\log T}. \tag{101}$$

Therefore, from (91), we know that the size of $\mathcal{N}_{s_i}$ can be bounded as

$$N_i \lesssim 13^d \log\left(\frac{1}{\sigma^2}\max_{s_i \leq t < s_{i+1}} \|\boldsymbol{\Lambda}_t^{-1/2}\boldsymbol{\eta}_t\|_2\right) \lesssim 13^d(\log d + \log\log T). \tag{102}$$

With this setup, we ensure that with probability $1 - \frac{1}{d\log^2 T}$, for any $s_i \leq t < s_{i+1}$, there exists $\boldsymbol{\lambda} \in \mathcal{N}_{s_i}$ such that

$$\left\|\boldsymbol{\Lambda}_t^{1/2}\boldsymbol{\lambda} - \frac{1}{\sigma^2}\boldsymbol{\Lambda}_t^{-1/2}\boldsymbol{\eta}_t\right\|_2 \leq \sqrt{2}\max_{1 \leq \|\boldsymbol{x}\|_2 \leq R}\left\|\boldsymbol{\Lambda}_{s_i}^{1/2}\boldsymbol{\lambda} - \boldsymbol{x}\right\|_2$$

$$\leq \frac{3}{4\sigma^2}\left\|\boldsymbol{\Lambda}_t^{-1/2}\boldsymbol{\eta}_t\right\|_2, \quad \text{whenever } \|\boldsymbol{\Lambda}_t^{-1/2}\boldsymbol{\eta}_t\|_2 \geq \sigma^2. \tag{103}$$

In this way, we ensure that with probability $1 - \frac{1}{\log T}$, for any $1 \leq t \leq T$, there exists some $\mathcal{N}_{s_i}$ and $\boldsymbol{\lambda} \in \mathcal{N}_{s_i}$ such that

$$\left\|\boldsymbol{\Lambda}_t^{1/2}\boldsymbol{\lambda} - \frac{1}{\sigma^2}\boldsymbol{\Lambda}_t^{-1/2}\boldsymbol{\eta}_t\right\|_2 \leq \frac{3}{4\sigma^2}\left\|\boldsymbol{\Lambda}_t^{-1/2}\boldsymbol{\eta}_t\right\|_2, \quad \text{whenever } \|\boldsymbol{\Lambda}_t^{-1/2}\boldsymbol{\eta}_t\|_2 \geq \sigma^2. \tag{104}$$

We guarantee that the weight $\overline{\gamma}_i$ that assigned on any points satisfies

$$\overline{\gamma}_i = \frac{1}{i(i+1)N_i} \gtrsim \frac{1}{13^d d^2 \log^2 T(\log d + \log\log T)} = \frac{1}{13^d\mathrm{poly}(d)\mathrm{poly}(\log T)}, \tag{105}$$

which concludes (84).

Proof of Claim (78). We note that for any $\boldsymbol{\lambda} \in \mathbb{R}^d$, the following inequality holds.

$$\mathbb{E}[M_T(\boldsymbol{\lambda}) \mid \mathcal{F}_t] = \mathbb{E}\left[M_t(\boldsymbol{\lambda})\exp\left(\boldsymbol{\lambda}^\top \boldsymbol{a}_T \epsilon_T - \frac{\sigma^2}{2}\boldsymbol{\lambda}^\top \boldsymbol{a}_T \boldsymbol{a}_T^\top \boldsymbol{\lambda}\right) \Big| \mathcal{F}_t\right]$$

$$= M_t(\boldsymbol{\lambda})\mathbb{E}\left[\exp\left(\boldsymbol{\lambda}^\top \boldsymbol{a}_T \epsilon_T\right) \mid \mathcal{F}_t\right]\exp\left(-\frac{\sigma^2}{2}\boldsymbol{\lambda}^\top \boldsymbol{a}_T \boldsymbol{a}_T^\top \boldsymbol{\lambda}\right)$$

$$\leq M_t(\boldsymbol{\lambda}), \tag{106}$$

where the last inequality holds because ϵ_t is σ sub-Gaussian random variable. As a result, we show that $M_t(\boldsymbol{\lambda})$ is a supermartingale.

Proof of Claim (96). Note that $\mathrm{tr}(\boldsymbol{\Lambda}_1) = d$ and $\mathrm{tr}(\boldsymbol{\Lambda}_t) = d + t - 1$. As a result, the determinant of $\det(\boldsymbol{\Lambda}_T)$ can be upper bounded as

$$\det(\boldsymbol{\Lambda}_T) \leq \left(\frac{d+T-1}{d}\right)^d, \tag{107}$$

and as a result, the number of switch times j can be bounded as

$$j \leq O\left[\log\left(\frac{\det(\boldsymbol{\Lambda}_T)}{\det(\boldsymbol{\Lambda}_1)}\right)\right] \leq O(d\log(d+T-1)) \leq O(d\log T). \tag{108}$$

For the second inequality, we note that $\mathrm{tr}(\boldsymbol{\Lambda}_{s_{i+1}}) \leq 2\mathrm{tr}(\boldsymbol{\Lambda}_{s_i})$. Therefore, $s_{i+1} + d - 1 \leq 2(s_i + d - 1)$. By induction, we note that $s_i + d - 1 \leq 2^{i-1}(s_1 + d - 1) = 2^{i-1}d$. As a result, we establish the inequality that

$$s_{i+1} - s_i \leq s_i + d - 1 \leq 2^{i-1}d. \tag{109}$$

D Proof of Theorem 3

We present the complete proof of Theorem 3 in this section. We begin with a high-level overview of the argument by phases in Appendix D.1, and then provide the detailed proofs of each phase from Appendix D.2–D.5.

For simplicity, throughout this section, we write with slight abuse of notation $\boldsymbol{\phi}(\boldsymbol{x}_t, \boldsymbol{a}_t) = \boldsymbol{a}_t$. While in general, the context can shift the mean and covariance of the action set, we can always transform it back to the unit ball, and so we utilize this for ease of notation. We also assume without loss of generality that the true signal $\boldsymbol{\theta}^\star = \boldsymbol{e}_1$ for our theoretical analysis, since the final result does not change up to a rotation. We also denote $\lambda_{t,1}, \ldots, \lambda_{t,d}$ as the eigenvalues of $\boldsymbol{\Lambda}_t$ ranked in decreasing order and $\boldsymbol{v}_{t,1}, \ldots, \boldsymbol{v}_{t,d}$ be the corresponding eigenvectors.

D.1 Key proof ideas

We summarize the high level proof ideas of each phase throughout the whole process.

Key analysis steps in Phase #1. The analysis of Phase #1 is divided into the following steps.

1. *Nontrivial mass on the bottom subspace.* In the initial phase, the ratio satisfies $\beta/\sqrt{\lambda_{t,d}} \gtrsim 1$. Consequently, $\mathrm{UCB}_t(\boldsymbol{a})$ is dominated by the exploration bonus rather than the estimated reward, favoring less–explored directions. Hence $\boldsymbol{a}_t$ necessarily places a constant fraction of its energy on the small–eigenvalue subspace (Lemma 3), i.e., on indices i with $\lambda_{t,i} \leq C\,\lambda_{t,d}$ for a fixed constant $C > 1$.

2. *Rank-one updates lift the bottom.* Each update is $\boldsymbol{a}_t \boldsymbol{a}_t^\top$, and as a result, the eigenvalue increments of $\boldsymbol{\Lambda}_t$ is approximated by the squared projections of $\boldsymbol{a}_t$ onto the eigenbasis (Lemma 5). Since $\boldsymbol{a}_t$ places a constant mass on the small-eigenvalue subspace, the small-eigenvalue block gains a uniformly positive aggregate amount each round.

3. *Lower bound for the minimum eigenvalue.* Distributing the persistent aggregate gain over at most d coordinates forces the minimum to rise at least linearly: the bottom block accrues a constant total increase each round, so its per–coordinate average grows by at least a constant multiple of $1/d$ per round, yielding $\lambda_{t,d} \gtrsim t/d$, for all t in the first stage when $\beta/\sqrt{\lambda_{t,d}} \gtrsim 1$, and in fact $\lambda_{t,d} \asymp t/d$ since it cannot exceed the average eigenvalue.

4. *Exit and eigengap.* Meanwhile, at some point when $\beta/\sqrt{\lambda_{t,d}} \lesssim 1$, the top eigenvalue grows strictly faster (by Theorem 2 and Lemma 2), while the rest are constrained by the mass-splitting above. Hence there exists t_1 such that $\beta/\sqrt{\lambda_{t_1,d}} \asymp 1$, an eigengap emerges, completing the first stage.

Key analysis steps in Phase #2. Set $\lambda_{t,d} = c_t \beta \sqrt{t}$ for $t \geq t_1$. It suffices to show $c_t \asymp d^{-1/2}$, which pins down $\lambda_{t,d} \asymp \beta\sqrt{t/d}$.

1. *Alignment of the top eigenvector.* A crude concentration bound gives $\|\boldsymbol{a}_t - \boldsymbol{\theta}^\star\|_2^2 \lesssim \beta/(c_t\sqrt{t})$. Since $\mathrm{tr}(\boldsymbol{\Lambda}_t) \approx t$, a Rayleigh–Ritz/Davis–Kahan argument yields $\|\boldsymbol{v}_{t,1} - \boldsymbol{\theta}^\star\|_2^2 \lesssim \beta/(\underline{c}_t\sqrt{t})$, with $\underline{c}_t = \min_{t_1 \leq s \leq t} c_s$. Here $\underline{c}_t$ is the historical normalized floor—the smallest value the normalized minimum eigenvalue has attained up to time t. The bound depends on $1/\underline{c}_t$ (rather than $1/c_t$) because it aggregates past rounds: the process cannot "forget" earlier times when the floor was lower and exploration bonuses were larger.

2. *Non-leading spectrum grows at the $\beta/\sqrt{t}$ scale.* Let $\overline{\lambda}_t$ be the mean of the non-leading eigenvalues. The one-step change $\overline{\lambda}_{t+1} - \overline{\lambda}_t$ can be upper bounded as $\overline{\lambda}_{t+1} - \overline{\lambda}_t = O(\beta/(d\underline{c}_t\sqrt{t}))$. Consequently, it is controlled by the alignment of $\boldsymbol{v}_{t,1}$ (better alignment means less spillover into non-leading directions), hence its dependence on the historical bottleneck $\underline{c}_t$. One-step changes can also be lower bounded as $\overline{\lambda}_{t+1} - \overline{\lambda}_t = \Omega(\beta/(dc_t\sqrt{t}))$, where c_t comes from the exploration of the UCB objective, which allocates nontrivial weight to under-explored directions and thus scales with the current floor c_t. As the result, the upper/lower envelopes match up to the ratio $c_t/\underline{c}_t$.

3. *Relative concentration on small eigenvalues.* For times with $c_t \leq \widetilde{c}$ (a fixed constant $\widetilde{c} > 0$) and $c_t/\underline{c_t} = O(1)$, we are in a regime where the per–coordinate exploration bonus $\beta w_{t,i}/\sqrt{\lambda_{t,i}}$ is comparable to the perturbation induced by the top direction (controlled by $\|\boldsymbol{v}_{t,1} - \boldsymbol{\theta}^\star\|$) for the non-leading coordinates. Consequently, the optimizer of UCB_t places a constant fraction of its non-leading mass on indices with small $\lambda_{t,i}$ (by Lemma 3). Repeating the rank-one update argument from Proposition 1 (cf. Lemma 5), this relative mass split transfers to growth: a fixed constant fraction of the total non-leading increment in each round is captured by the small-eigenvalue block.

4. *Forcing up the minimum and closing the loop.* Because a fixed fraction of the non-leading increment lands on the bottom block each round, and that $\overline{\lambda}_t$ enjoys a per-step lower bound shown in Step 2, we can show that $\lambda_{t,d}$ grows at least at the rate of $\Omega(\beta/(dc_t\sqrt{t}))$ in this regime, as a result, $c_t \gtrsim 1/(dc_t)$, leading to a lower bound $\lambda_{t,d} \gtrsim \beta\sqrt{t}/\sqrt{d}$, when $t \geq t_1$. Conversely, $\lambda_{t,d}$ cannot exceed the non-leading average, yielding the matching upper bound and hence $\lambda_{t,d} \asymp \beta\sqrt{t/d}$. Plugging this back into the alignment bound gives $\|\boldsymbol{v}_{t,1} - \boldsymbol{\theta}^\star\|_2^2 \lesssim \beta^2/\lambda_{t,d}$, which completes the argument.

Key analysis steps in Phase #3. We present the main steps of the proof of Phase #3 as follows.

1. We track how the leading eigenvector $\boldsymbol{v}_{t,1}$ evolves as the sample size increases from t to $t+1$. Using that $\boldsymbol{a}_t$ optimizes the UCB objective, the new sample induces a rank-one perturbation linking $(\boldsymbol{v}_{t,1}, \widehat{\boldsymbol{\theta}}_t)$ to $\boldsymbol{v}_{t+1,1}$:

$$\boldsymbol{v}_{t+1,1} = \boldsymbol{v}_{t,1} + \frac{\widehat{\boldsymbol{\theta}}_t - \boldsymbol{v}_{t,1}}{t} + \boldsymbol{\zeta}_t.$$

The fluctuation term $\boldsymbol{\zeta}_t$ has, in the worst case, only higher-order adverse effect on alignment with $\widehat{\boldsymbol{\theta}}_t$, whereas any component that improves alignment may be non-negligible. Consequently, the update makes $\boldsymbol{v}_{t+1,1}$ closer to $\widehat{\boldsymbol{\theta}}_t$ than $\boldsymbol{v}_{t,1}$. As a result, $\|\boldsymbol{v}_{t,1} - \boldsymbol{\theta}^\star\|_2$ decreases until it is of the same order as $\|\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^\star\|_2$.

2. Leveraging upon the above update, we study how error propogates with time step t and obtain the following

$$\|\boldsymbol{v}_{t+1,1} - \boldsymbol{\theta}^\star\|_2^2 \leq \left(\left(1 - \frac{1}{t}\right)\|\boldsymbol{v}_{t,1} - \boldsymbol{\theta}^\star\|_2 + \widetilde{O}(t^{-5/4})\right)^2 + \widetilde{O}(t^{-2}), \tag{110}$$

so the error contracts by roughly $(1 - 1/t)$ with faster-vanishing additive terms. Combining this with the Phase #2 initialization, there exists $t_2 = O(\beta^8/(\sigma^6 d^2))$ such that for all $t \geq t_2$, $\boldsymbol{v}_{t,1}$ attains the desired concentration around $\boldsymbol{\theta}^\star$, matching the concentration order of $\widehat{\boldsymbol{\theta}}_t$.

Key analysis steps in Phase #4. We outline the key steps in Phase #4 as follows.

1. *Precise decomposition of $\boldsymbol{a}_t$.* Using the refined concentration of the top eigenvector (the high-probability bound on $\|\boldsymbol{v}_{t,1} - \widehat{\boldsymbol{\theta}}_t\|_2$), we sharpen Lemma 2 and make explicit how $\boldsymbol{a}_t$ splits between the leading direction and its orthogonal complement:

   - *Non-leading directions.* The non-leading coordinates of $\boldsymbol{w}_t$ (defined in (35)) satisfy

   $$\sum_{i=2}^{d} w_{t,i}^2 \left(1 - \frac{\lambda_{t,d}}{\lambda_{t,i}}\right) = o(1).$$

   This equation implies that $\boldsymbol{w}_t$, and hence $\boldsymbol{a}_t$, concentrates most of its non-leading mass on directions whose eigenvalues are close to $\lambda_{t,d}$.

   - *Mass off the top.* The portion of $\boldsymbol{a}_t$ lying outside the top direction admits the precise characterization

   $$\sum_{i=2}^{d} \kappa_{t,i}^2 = \frac{\beta^2}{\lambda_{t,d}} \left(1 - \frac{\lambda_{t,d}^2}{\beta^2 t} + o(1)\right),$$

as long as $\lambda_{t,d} \leq \beta\sqrt{t}$. Therefore, we can precisely characterize the growth speed of non-leading eigenvalues under this regime.

2. *Preliminary growth-rate control for $\lambda_{t,d}$.* We prove that there exists $t_2' = O(t_2)$ such that for all $t \geq t_2'$, $\lambda_{t,d} \leq c\beta\sqrt{t}$ for some constant $c < 1$. This bound is essential: as shown in the previous step, the growth of the non-leading eigenvalues can only be characterized sharply when $\lambda_{t,d}$ is smaller than $\beta\sqrt{t}$. We need this requirement to establish further fine-grained arguments.

3. *Limiting the projection of $\boldsymbol{a}_t$ onto eigenspaces with large eigenvalues.* Using the precise decomposition from Step 1, define the set of large eigenvalues as those exceeding $(1 + \delta)\lambda_{t,d}$ with $\delta = o(1)$. Then $\boldsymbol{a}_t$ allocates at most an $O(1/d)$ fraction of its non-leading mass to these large eigenvalues.

4. *Controlling the growth of large non-leading eigenvalues.* We show that eigenvalues above the $(1+d\delta)\lambda_{t,d}$ threshold do not grow faster than $\overline{\lambda}_t$. Consequently, the non-leading eigenvalues concentrate, differing in magnitude by at most a $(1 + o(1))$ factor.

5. *Precise characterization of non-leading eigenvalues.* By previous results, the non-leading eigenvalues coalesce: for $t \geq t_4$, $\lambda_{t,2} \approx \cdots \approx \lambda_{t,d}$. It is therefore natural to track their common level via the average $\overline{\lambda}_t := \frac{1}{d-1}\sum_{i=2}^{d} \lambda_{t,i}$. For $t \geq t_4$, $\overline{\lambda}_t$ evolves according to

$$\overline{\lambda}_T = \overline{\lambda}_t + \frac{1}{d-1}\frac{\beta^2}{\overline{\lambda}_t}\left(1 - \frac{\overline{\lambda}_t^2}{\beta^2 t} + o(1)\right).$$

This recursion makes the growth of $\overline{\lambda}_t$ explicit: to leading order it increases at rate $\beta^2/((d-1)\overline{\lambda}_t)$, with a vanishing correction of order $\overline{\lambda}_t/(\beta^2 t)$. In turn, it delivers a precise large-t asymptotic for $\overline{\lambda}_t$ together with its first-order correction. This constructs our desired result.

D.2   Analysis of Phase #1 (proof of Proposition 1)

We aim to show that whenever $\beta/\sqrt{\lambda_{t,d}} \geq C'$, there exists constant $C = C(C')$ such that

$$\lambda_{t,d} \geq C \cdot \frac{t}{d}.$$

Step 1: lower bound the projection of a_t on eigenspaces with "low eigenvalues."   At the beginning, the following conditions are satisfied:

$$\frac{\beta}{\sqrt{\lambda_{t,d}}} \geq C', \qquad \sum_{i=1}^{d} \nu_{t,i}^2 = 1,$$

for some constant $C' > 0$. The first inequality guarantees that the effective signal-to-noise ratio remains bounded away from zero, while the second condition normalizes the direction vector ν_t.

Under these assumptions, the optimization problem (35) falls within the scope of Lemma 3. By direct application of this lemma, there exist absolute constants $C_1 = C_1(C')$ and $C_2 \in (0, 1]$ such that, if we define

$$k_t = \max\{i : \lambda_{t,i} > C_1\lambda_{t,d}\},$$

then the coefficients $\kappa_{t,i}$ in the expansion of $\boldsymbol{a}_t$ must satisfy

$$\sum_{i=k_t+1}^{d} \kappa_{t,i}^2 \geq C_2. \tag{111}$$

This inequality formalizes the intuition that a nontrivial fraction of the action vector necessarily lies in the less dominant eigenspaces—those associated with eigenvalues not substantially larger than $\lambda_{t,d}$. In other words, $\boldsymbol{a}_t$ cannot concentrate exclusively on the top eigen-directions; a uniformly positive share of its energy always projects onto the "low-eigenvalue" space.

Step 2: control the growth of "large eigenvalues". Next, define the set of "large eigenvalues" to be those exceeding $2C_1\lambda_{t,d}$. Let

$$k_t = \max\{j : \lambda_{t,j} > 2C_1\lambda_{t,d}\}.$$

From Lemma 5, the updated eigenvalues $\lambda_{t+1,1}, \ldots, \lambda_{t+1,d}$ are precisely the roots of

$$f(\lambda) = 1 + \sum_{i=1}^{d} \frac{\kappa_{t,i}^2}{\lambda_{t,i} - \lambda}.$$

To isolate the contribution of the "large" coordinates, we define the auxiliary function

$$f_1(\lambda) = 1 + \sum_{i=1}^{k_t} \frac{\kappa_{t,i}^2}{\lambda_{t,i} - \lambda}.$$

For any $i \leq k_t$, one has

$$f(\lambda_{t+1,i}) = f_1(\lambda_{t+1,i}) + \sum_{i=k_t+1}^{d} \frac{\kappa_{t,i}^2}{\lambda_{t,i} - \lambda_{t+1,i}} \geq f_1(\lambda_{t+1,i}) - \frac{1}{C_1\lambda_{t,d}} \sum_{i=k_t+1}^{d} \kappa_{t,i}^2, \tag{112}$$

which implies

$$f_1(\lambda_{t+1,i}) \leq \frac{1}{C_1} \sum_{i=k_t+1}^{d} \kappa_{t,i}^2. \tag{113}$$

Let $\widetilde{\lambda}_{t,1}, \ldots, \widetilde{\lambda}_{t,k_t}$ be the solutions of

$$f_1(\lambda) = \frac{1}{C_1} \sum_{i=k_t+1}^{d} \kappa_{t,i}^2.$$

Equivalently, these roots satisfy

$$\begin{aligned}
f_1(\lambda) - \frac{1}{C_1} \sum_{i=k_t+1}^{d} \kappa_{t,i}^2 &= 1 - \frac{1}{C_1} \sum_{i=k_t+1}^{d} \kappa_{t,i}^2 + \sum_{i=1}^{k_t} \frac{\kappa_{t,i}^2}{\lambda_{t,i} - \lambda} \\
&= \frac{\left(1 - \frac{1}{C_1}\sum_{i=k_t+1}^{d}\kappa_{t,i}^2\right) \cdot \prod_{i=1}^{k_t}(\lambda_{t,i} - \lambda) + \sum_{i=1}^{k_t}\kappa_{t,i}^2 \prod_{j\neq i}^{k_t}(\lambda_{t,j} - \lambda)}{\prod_{i=1}^{k_t}(\lambda_{t,i} - \lambda)}.
\end{aligned} \tag{114}$$

By examining the coefficients of the characteristic polynomial, we identify the leading coefficients:

$$m_1 = (-1)^{k_t}\left(1 - \frac{1}{C_1}\sum_{i=k_t+1}^{d}\kappa_{t,i}^2\right),$$

and

$$m_2 = (-1)^{k_t-1}\left[\left(1 - \frac{1}{C_1}\sum_{i=k_t+1}^{d}\kappa_{t,i}^2\right) \cdot \sum_{i=1}^{k_t}\lambda_i + \sum_{i=1}^{k_t}\kappa_{t,i}^2\right].$$

Hence, the sum of the auxiliary roots satisfies

$$\sum_{i=1}^{k_t}\widetilde{\lambda}_{t+1,i} = -\frac{m_2}{m_1} = \sum_{i=1}^{k_t}\lambda_i + \frac{\sum_{i=1}^{k_t}\kappa_{t,i}^2}{1 - \frac{1}{C_1}\sum_{i=k_t+1}^{d}\kappa_{t,i}^2} \leq \sum_{i=1}^{k_t}\lambda_i + \frac{1 - C_2}{1 - \frac{C_2}{C_1}}. \tag{115}$$

Combining this with (113), we deduce that for all $i \leq j_t$,

$$\lambda_{t+1,i} \leq \widetilde{\lambda}_{t+1,i}.$$

Therefore, the total mass of the top j_t eigenvalues after the update is bounded as

$$\sum_{i=1}^{j_t} \lambda_{t+1,i} \leq \sum_{i=1}^{j_t} \widetilde{\lambda}_{t+1,i} = \sum_{i=1}^{j_t} \lambda_{t,i} + \sum_{i=1}^{j_t} (\widetilde{\lambda}_{t+1,i} - \lambda_{t,i}) \leq \sum_{i=1}^{j_t} \lambda_{t,i} + \frac{1 - C_2}{1 - \frac{C_2}{C_1}}$$

$$\leq \sum_{i=1}^{j_t} \lambda_{t,i} + \frac{1 - C_2}{1 - \frac{C_2}{2}} \leq \sum_{i=1}^{j_t} \lambda_{t,i} + 1 - \frac{C_2}{2}. \tag{116}$$

Consequently, the remaining eigenvalues necessarily gain at least a fixed amount of mass:

$$\sum_{i=j_t+1}^{d} \lambda_{t+1,i} \geq \sum_{i=j_t+1}^{d} \lambda_{t,i} + \frac{C_2}{2}. \tag{117}$$

Step 3: lower bound on the smallest eigenvalue. We now use the results from the previous parts to establish a quantitative lower bound for $\lambda_{t,d}$. Recall that in Step 2 we showed that for the set

$$\mathcal{L}_t = \{\, i : \lambda_{t,i} \leq 2C_1\lambda_{t,d} \,\},$$

the projection of $\boldsymbol{a}_t$ satisfies

$$\sum_{i \in \mathcal{L}_t} \kappa_{t,i}^2 \geq \frac{C_2}{2}.$$

To track the cumulative effect across time, we extend this notation. For each $t \leq t'$, define

$$\overline{\mathcal{L}}_{t,t'} = \{\, i : \lambda_{t,i} \leq 2C_1\lambda_{t',d} \,\}.$$

Clearly, $\mathcal{L}_t \subseteq \overline{\mathcal{L}}_{t,t'}$. Next, define A_t to be the total gain accumulated by the eigenvalues in $\mathcal{L}_s$ up to time t:

$$A_t = \sum_{s=0}^{t-1} \sum_{i \in \mathcal{L}_s} \left(\lambda_{s+1,i} - \lambda_{s,i} \right). \tag{118}$$

From Step 2, each summand contributes at least $C_2/2$, which yields the lower bound

$$A_t \geq \sum_{s=0}^{t-1} \frac{C_2}{2} = \frac{C_2 t}{2}.$$

On the other hand, we can upper bound A_t using the enlarged sets $\overline{\mathcal{L}}_{s,t}$:

$$A_t \leq \sum_{s=0}^{t-1} \sum_{i \in \overline{\mathcal{L}}_{s,t}} \left(\lambda_{s+1,i} - \lambda_{s,i} \right)$$

$$= \sum_{s=0}^{t-1} \sum_{i=1}^{d} \left(\lambda_{s+1,i} - \lambda_{s,i} \right) \mathbf{1}\{\lambda_{s,i} \leq 2C_1\lambda_{t,d}\}$$

$$= \sum_{i=1}^{d} \sum_{s=0}^{t-1} \left(\lambda_{s+1,i} - \lambda_{s,i} \right) \mathbf{1}\{\lambda_{s,i} \leq 2C_1\lambda_{t,d}\}. \tag{119}$$

Since each $\lambda_{s,i}$ is monotone increasing, the inner summation is a telescoping sum. Moreover, whenever $\lambda_{s,i} \leq 2C_1\lambda_{t,d}$ we have

$$\lambda_{s+1,i} \leq \lambda_{s,i} + 1 \leq 2C_1\lambda_{t,d} + 1.$$

Thus the total increase of each eigenvalue is bounded by $2C_1\lambda_{t,d}$, recalling that $\boldsymbol{\Lambda}_0 = \boldsymbol{I}_d$ (so $\lambda_{0,i} = 1$ for all i). Therefore,

$$A_t \leq \sum_{i=1}^{d} 2C_1\lambda_{t,d} \leq 2C_1 d\, \lambda_{t,d}. \tag{120}$$

Combining the lower bound (118) with the upper bound (120), we obtain

$$\lambda_{t,d} \;\geq\; \frac{C_2 t}{4 C_1 d}.$$

Therefore, we have shown that whenever $\beta/\sqrt{\lambda_{t,d}} \geq 1/2$, there exists a universal constant $C > 0$ such that

$$\lambda_{t,d} \;\geq\; C \cdot \frac{t}{d}.$$

Step 4: growth of the leading eigenvalue. Define the stopping time

$$t_1' = \min\left\{ t : \frac{\beta}{\sqrt{\lambda_{t,d}}} \leq C' \right\},$$

for some constant C'. From the conclusion of Step 3, it follows immediately that $t_1' = \Theta(\beta^2 d)$. For any $t \geq t_1'$, with probability $1 - 1/T$, the estimation error can be bounded as

$$\left\| \boldsymbol{a}_t - \boldsymbol{\theta}^\star \right\|_2 \leq \left\| \boldsymbol{a}_t - \widehat{\boldsymbol{\theta}}_t \right\|_2 + \left\| \widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^\star \right\|_2 \lesssim \frac{\beta}{\sqrt{\lambda_{t,d}}} + \frac{\beta}{\sqrt{\lambda_{t,d}}} \leq \frac{3}{5}, \tag{121}$$

when C' is set small enough, where the second inequality holds from the confidence set in (2), where we set $\delta = 1/T$. As a consequence, for all $t > t_1'$ we obtain the correlation guarantee

$$\langle \boldsymbol{a}_t, \boldsymbol{\theta}^\star \rangle \;\geq\; \sqrt{1 - \left(\frac{3}{5}\right)^2} \;=\; \frac{4}{5}.$$

Hence, setting $t_1 = 6 t_1'$, we deduce

$$\boldsymbol{\theta}^{\star T} \boldsymbol{\Lambda}_{t_1} \boldsymbol{\theta}^\star = \sum_{s=t_1'+1}^{t_1} \langle \boldsymbol{\theta}^\star, \boldsymbol{a}_t \rangle^2 \geq \sum_{s=t_1'+1}^{t_1} \left(\frac{4}{5}\right)^2 \geq \frac{8}{15} t_1, \tag{122}$$

which implies

$$\lambda_{t_1,1} \;\geq\; \frac{8}{15} t_1.$$

Therefore, the leading eigenvalue separates from the rest by

$$\lambda_{t_1,1} - \lambda_{t_1,2} \;\geq\; \frac{t_1}{15}.$$

D.3 Analysis of Phase #2 (proof of Proposition 2)

We reparameterize $\lambda_{t,d}$ as follows: define c_t such that

$$\lambda_{t,d} = c_t \beta \sqrt{t}. \tag{123}$$

We further introduce the notation

$$\underline{c}_t = \min_{t_1 \leq s \leq t} c_s,$$

and define $\overline{\lambda}_t$ as the empirical average of the non-leading eigenvalues:

$$\overline{\lambda}_t = \frac{1}{d-1} \sum_{i=2}^{d} \lambda_{t,i}. \tag{124}$$

The key to establishing the desired result is to show that there exists a constant $c^\star > 0$ such that

$$0 < c^\star < \underline{c}_t \qquad \text{for all } t > t_1.$$

39

Step 1: characterize the distance between $\boldsymbol{v}_{t,1}$ and $\boldsymbol{\theta}^\star$. We begin with the standard confidence bound on the estimation error:

$$\|\boldsymbol{a}_t - \widehat{\boldsymbol{\theta}}_t\|_2 \lesssim \frac{\beta}{\sqrt{\lambda_{t,d}}}. \tag{125}$$

This shows that the action $\boldsymbol{a}_t$ chosen at time t is close to the current estimate $\widehat{\boldsymbol{\theta}}_t$, with the error shrinking as the smallest eigenvalue $\lambda_{t,d}$ grows. By the triangle inequality, this further implies closeness to the true parameter: with probability $1 - 1/T$,

$$\|\boldsymbol{a}_t - \boldsymbol{\theta}^\star\|_2 \lesssim \frac{\beta}{\sqrt{\lambda_{t,d}}} + \frac{\beta}{\sqrt{\lambda_{t,d}}} \lesssim \frac{\beta}{\sqrt{\lambda_{t,d}}}, \tag{126}$$

where the first inequality follow directly from (121). Intuitively, this condition ensures that the statistical noise is dominated by the exploration parameter β, keeping $\boldsymbol{a}_t$ well aligned with $\boldsymbol{\theta}^\star$. Next, consider the magnitude of the "transformed signal" $\boldsymbol{\Lambda}_t^{1/2}\boldsymbol{\theta}^\star$. Its squared norm is the cumulative signal energy collected along $\boldsymbol{\theta}^\star$:

$$\begin{aligned}
\left\|\boldsymbol{\Lambda}_t^{1/2}\boldsymbol{\theta}^\star\right\|_2 &= \sqrt{\boldsymbol{\theta}^{\star T}\boldsymbol{\Lambda}_t\boldsymbol{\theta}^\star} = \sqrt{\sum_{s=1}^t \langle \boldsymbol{a}_s, \boldsymbol{\theta}^\star \rangle^2} \\
&= \sqrt{\sum_{s=1}^t \left(1 - O\left(\frac{\beta^2}{\lambda_{s,d}}\right)\right)} = \sqrt{\sum_{s=1}^t \left(1 - O\left(\frac{\beta}{c_s\sqrt{s}}\right)\right)} \\
&= \sqrt{t - O\left(\frac{\beta\sqrt{t}}{c_t}\right)}.
\end{aligned} \tag{127}$$

Thus, the information collected in the direction of $\boldsymbol{\theta}^\star$ grows like $\sqrt{t}$, with only a mild correction due to imperfect exploration. By comparison, the transformed leading eigenvector has energy bounded by the trace:

$$\left\|\boldsymbol{\Lambda}_t^{1/2}\boldsymbol{v}_{t,1}\right\|_2 = \sqrt{\boldsymbol{v}_{t,1}^T\boldsymbol{\Lambda}_t\boldsymbol{v}_{t,1}} \le t + d. \tag{128}$$

This is a crude upper bound, but sufficient for our purposes. We now turn to the eigenvalue structure. The largest eigenvalue satisfies

$$\lambda_{t,1} \ge \left\|\boldsymbol{\Lambda}_t^{1/2}\boldsymbol{\theta}^\star\right\|_2^2 = t - O\left(\frac{\beta\sqrt{t}}{c_t}\right). \tag{129}$$

To relate $\boldsymbol{v}_{t,1}$ to $\boldsymbol{\theta}^\star$, expand $\boldsymbol{\theta}^\star$ in the eigenbasis:

$$\begin{aligned}
\left\|\boldsymbol{\Lambda}_t^{1/2}\boldsymbol{\theta}^\star\right\|_2^2 &= \sum_{i=1}^d \lambda_{t,i} \langle \boldsymbol{\theta}^\star, \boldsymbol{v}_{t,i} \rangle^2 \\
&\le \lambda_{t,1}\langle \boldsymbol{\theta}^\star, \boldsymbol{v}_{t,1} \rangle^2 + \lambda_{t,2}\left(1 - \langle \boldsymbol{\theta}^\star, \boldsymbol{v}_{t,1} \rangle^2\right) \\
&\le \lambda_{t,2} + (\lambda_{t,1} - \lambda_{t,2}) \cdot \langle \boldsymbol{\theta}^\star, \boldsymbol{v}_{t,1} \rangle^2.
\end{aligned} \tag{130}$$

Since $\mathrm{tr}(\boldsymbol{\Lambda}_t) = t + d$, the contribution from the smaller eigenvalues is limited:

$$\sum_{i \ge 2} \lambda_{t,i} = t + d - \lambda_{t,1} = O\left(\frac{\beta\sqrt{t}}{c_t}\right). \tag{131}$$

Thus, $\lambda_{t,2} \le O\left(\frac{\beta\sqrt{t}}{c_t}\right)$. It is also straightforward that $\lambda_{t,1} - \lambda_{t,2} \le t$. Plugging this into the expansion gives

$$\left\|\boldsymbol{\Lambda}_t^{1/2}\boldsymbol{\theta}^\star\right\|_2^2 \le O\left(\frac{\beta\sqrt{t}}{c_t}\right) + t \cdot \langle \boldsymbol{\theta}^\star, \boldsymbol{v}_{t,1} \rangle^2. \tag{132}$$

On the other hand, we already have the lower bound

$$\left\| \mathbf{\Lambda}_t^{1/2} \boldsymbol{\theta}^\star \right\|_2^2 \geq t - O\left(\frac{\beta\sqrt{t}}{\underline{c}_t} \right).$$

Together, these inequalities imply

$$t - O\left(\frac{\beta\sqrt{t}}{\underline{c}_t} \right) \;\leq\; t \cdot \langle \boldsymbol{\theta}^\star, \boldsymbol{v}_{t,1} \rangle^2 + O\left(\frac{\beta\sqrt{t}}{\underline{c}_t} \right). \tag{133}$$

Rearranging yields

$$\langle \boldsymbol{\theta}^\star, \boldsymbol{v}_{t,1} \rangle^2 \geq 1 - O\left(\frac{\beta}{\underline{c}_t \sqrt{t}} \right). \tag{134}$$

Finally, for unit vectors it is well known that

$$\langle \boldsymbol{\theta}^\star, \boldsymbol{v}_{t,1} \rangle^2 = 1 - \Omega\big(\| \boldsymbol{v}_{t,1} - \boldsymbol{\theta}^\star \|_2^2 \big).$$

Hence the above inequality translates into the key control

$$\| \boldsymbol{v}_{t,1} - \boldsymbol{\theta}^\star \|_2^2 = O\left(\frac{\beta}{\underline{c}_t \sqrt{t}} \right). \tag{135}$$

Step 2: upper and lower bound the growth speed of non-leading eigenvalues. We next analyze how the non-leading eigenvalues evolve over time. Establishing both upper and lower bounds on their growth is important: the upper bound ensures that they do not accumulate too much mass relative to the leading eigenvalue, while the lower bound guarantees that they still grow at a sufficient rate to prevent degeneration.

Upper bound. By Lemma 2, the decomposition coefficients satisfy

$$\nu_{t,1} = 1 - O\left(\frac{\beta}{\underline{c}_t \sqrt{t}} \right), \qquad \kappa_{t,1} = 1 - O\left(\frac{\beta}{\underline{c}_t \sqrt{t}} \right). \tag{136}$$

This shows that the majority of the mass concentrates in the leading eigen-direction, with only a small error proportional to $\frac{\beta}{\underline{c}_t \sqrt{t}}$. Applying Lemma 6, the leading eigenvalue evolves as

$$\lambda_{t+1,1} = \lambda_{t,1} + \kappa_{t,1}^2 + O\big(t^{-1} \big). \tag{137}$$

Because $\kappa_{t,1}^2 \approx 1$, this increment essentially captures the rate at which the leading eigenvalue dominates. Consequently, the mean of the non-leading eigenvalues evolves according to

$$\overline{\lambda}_T = \overline{\lambda}_t + \frac{1}{d-1} \cdot O\left(\frac{\beta}{\underline{c}_t \sqrt{t}} \right). \tag{138}$$

Since $\underline{c}_t$ is non-increasing, telescoping this recursion yields the global upper bound

$$\overline{\lambda}_t = O\left(\frac{\beta\sqrt{t}}{d\underline{c}_t} \right). \tag{139}$$

In other words, the mean non-leading eigenvalue cannot grow faster than $\sqrt{t}$, up to a factor depending on β and the stability term $\underline{c}_t$.

Lower bound. To complement the above, we construct a lower bound by considering the auxiliary optimization problem

$$\widetilde{\boldsymbol{w}}_t = \arg \max_{\| \boldsymbol{w} \|_2 = 1} \left(1 + \frac{\beta w_1}{\sqrt{\lambda_{t,1}}} \right)^2 + \sum_{i=2}^d \frac{\beta^2 w_i^2}{\lambda_{t,i}}. \tag{140}$$

41

This problem identifies the direction that maximizes the quadratic growth contribution, balancing the leading component with those from the non-leading directions. Expanding the expression shows that this is equivalent to maximizing

$$\max_{\|\boldsymbol{w}\|_2=1} \frac{2\beta w_{t,1}}{\sqrt{\lambda_{t,1}}} - \beta^2 w_{t,1}^2 \left(\frac{1}{\lambda_{t,d}} - \frac{1}{\lambda_{t,1}}\right). \tag{141}$$

Thus, the optimal weight on the leading coordinate, denoted $\tilde{w}_{t,1}^\star$, must satisfy

$$\tilde{w}_{t,1}^\star = \min\left(1, \frac{1}{\beta\sqrt{\lambda_{t,1}}} \left(\frac{1}{\lambda_{t,d}} - \frac{1}{\lambda_{t,1}}\right)^{-1}\right). \tag{142}$$

In Phase 2, since $\lambda_{t,1} = \Omega(t)$ and the eigengap obeys $\lambda_{t,1} - \lambda_{t,d} = \Omega(\lambda_{t,1})$, this reduces to

$$\tilde{w}_{t,1}^\star = \min\left(1, O\left(\frac{\lambda_{t,d}}{\beta\sqrt{t}}\right)\right) \leq O\left(\frac{\lambda_{t,d}}{\beta\sqrt{t}}\right). \tag{143}$$

Hence the mass allocated to the leading coordinate is negligible whenever $\lambda_{t,d}$ is small, meaning most weight shifts to non-leading directions. Therefore, the contribution from the non-leading coordinates satisfies

$$\sum_{i=2}^{d} \frac{\beta^2 (\tilde{w}_{t,i}^\star)^2}{\lambda_{t,i}} = \frac{\beta^2}{\lambda_{t,d}} \left(1 - O\left(\frac{\lambda_{t,d}^2}{\beta^2 t}\right)\right). \tag{144}$$

By Lemma 4, this translates to the inequality

$$\sum_{i=2}^{d} \left(\nu_{t,i} + \frac{\beta w_{t,i}^\star}{\sqrt{\lambda_{t,i}}}\right)^2 \geq \frac{\beta^2 \left(1 - (w_{t,1}^\star)^2\right)}{\lambda_{t,d}} \geq \frac{\beta^2 \left(1 - (\tilde{w}_{t,1}^\star)^2\right)}{\lambda_{t,d}}. \tag{145}$$

Meanwhile, the total contribution is bounded by

$$\sum_{i=1}^{d} \left(\nu_{t,i} + \frac{\beta w_{t,i}^\star}{\sqrt{\lambda_{t,i}}}\right)^2 \leq 1 + \frac{\beta}{\sqrt{\lambda_{t,d}}} = O(1), \tag{146}$$

so it follows that

$$\sum_{i=2}^{d} \kappa_{t,i}^2 \gtrsim \frac{\beta^2}{\lambda_{t,d}} \left(1 - O\left(\frac{\lambda_{t,d}^2}{\beta^2 t}\right)\right). \tag{147}$$

Finally, recalling that the average of the non-leading eigenvalues evolves as

$$\overline{\lambda}_{t+1} = \overline{\lambda}_t + \frac{\sum_{i=2}^{d} \kappa_{t,i}^2}{d-1} + O\left(t^{-1}\right), \tag{148}$$

we conclude that whenever $\lambda_{t,d} \leq \tilde{c}\beta\sqrt{t}$ (for some universal constant $\tilde{c}$), the recursion satisfies

$$\overline{\lambda}_{t+1} = \overline{\lambda}_t + \frac{1}{d-1} \cdot \Omega\left(\frac{\beta^2}{\lambda_{t,d}}\right) = \overline{\lambda}_t + \frac{1}{d-1} \cdot \Omega\left(\frac{\beta}{c_t\sqrt{t}}\right). \tag{149}$$

Step 3: lower bound the growth of "small eigenvalues". We now aim to establish a lower bound on the growth rate of the set of "small eigenvalues" in the regime where t belongs to the set

$$\mathcal{S} = \left\{t : c_t \leq \tilde{c}, \frac{c_t}{\underline{c}_t} \leq 2\right\},$$

that is, the regime in which the smallest eigenvalue risks dropping below the desired growth rate and has already crossed a critical threshold. Our proof strategy parallels that of Phase #1, but with a crucial

42

modification. In this phase, we bound the growth rate of the small eigenvalues from below by a constant multiple of the growth rate of the mean of the non-leading eigenvalues, rather than by the mean of all eigenvalues. This adjustment is necessary because the leading eigenvalue $\lambda_{t,1}$ grows too rapidly to serve as a meaningful reference point in this regime.

To proceed, consider the optimization problem

$$
\begin{aligned}
\max_{\|\boldsymbol{w}_t\|_2=1} & \sum_{i=1}^{d} \left(\nu_{t,i} + \frac{\beta w_{t,i}}{\sqrt{\lambda_{t,i}}} \right)^2 \\
= \max_{w_1} & \max_{\|\boldsymbol{w}_{-1}\|_2=\sqrt{1-w_1^2}} \left[\left(\nu_{t,1} + \frac{\beta w_{t,1}}{\sqrt{\lambda_{t,1}}} \right)^2 + \sum_{i=2}^{d} \left(\nu_{t,i} + \frac{\beta w_{t,i}}{\sqrt{\lambda_{t,i}}} \right)^2 \right] \\
= \max_{w_1} & \left(\nu_{t,1} + \frac{\beta w_{t,1}}{\sqrt{\lambda_{t,1}}} \right)^2 + \max_{\|\boldsymbol{w}_{-1}\|_2=\sqrt{1-w_1^2}} \sum_{i=2}^{d} \left(\nu_{t,i} + \frac{\beta w_{t,i}}{\sqrt{\lambda_{t,i}}} \right)^2.
\end{aligned}
\tag{150}
$$

Thus, to characterize the contribution from $(w_{t,2},\ldots,w_{t,d})$, it suffices to analyze the sub-optimization problem

$$
\max_{\|\boldsymbol{w}_{-1}\|_2=\sqrt{1-w_1^2}} \sum_{i=2}^{d} \left(\nu_{t,i} + \frac{\beta w_{t,i}}{\sqrt{\lambda_{t,i}}} \right)^2.
\tag{151}
$$

Since $c_t/\underline{c}_t \leq 2$, it follows that

$$
\nu_{t,1} = 1 - O\left(\frac{\beta}{\underline{c}_t\sqrt{t}} \right) = 1 - O\left(\frac{\beta}{c_t\sqrt{t}} \right) = 1 - O\left(\frac{\beta^2}{\lambda_{t,d}} \right),
\tag{152}
$$

where the last step uses $\lambda_{t,d} = c_t\beta\sqrt{t}$. Consequently,

$$
\sqrt{\sum_{i=2}^{d} \nu_{t,i}^2} = O\left(\frac{\beta}{\sqrt{\lambda_{t,d}}} \right).
\tag{153}
$$

Returning to the optimization, we obtain

$$
\begin{aligned}
& \left(\nu_{t,1} + \frac{\beta w_{t,1}}{\sqrt{\lambda_{t,1}}} \right)^2 + \max_{\|\boldsymbol{w}_{-1}\|_2=\sqrt{1-w_1^2}} \sum_{i=2}^{d} \left(\nu_{t,i} + \frac{\beta w_{t,i}}{\sqrt{\lambda_{t,i}}} \right)^2 \\
\leq & \left(\nu_{t,1} + \frac{\beta w_{t,1}}{\sqrt{\lambda_{t,1}}} \right)^2 + \left(\sqrt{\sum_{i=2}^{d} \nu_{t,i}^2} + \frac{\beta}{\sqrt{\lambda_{t,d}}}\sqrt{\sum_{i=2}^{d} w_{t,i}^2} \right)^2 \\
\leq & 1 + O\left(\frac{\beta}{\sqrt{\lambda_{t,1}}} \right) + O\left(\frac{\beta^2}{\lambda_{t,d}} \right) \cdot \sqrt{\sum_{i=2}^{d} w_{t,i}^2}.
\end{aligned}
\tag{154}
$$

At the same time, the optimization also admits a simple lower bound:

$$
\max_{\|\boldsymbol{w}_t\|_2=1} \sum_{i=1}^{d} \left(\nu_{t,i} + \frac{\beta w_{t,i}}{\sqrt{\lambda_{t,i}}} \right)^2 \geq 1 + \frac{\beta^2}{\lambda_{t,d}}.
\tag{155}
$$

Since $\lambda_{t,1} \geq t/2$ and $\lambda_{t,d} = c_t\beta\sqrt{t}$, we deduce that when c_t is sufficiently small (which can be achieved by choosing a sufficiently small constant $\widetilde{c}$), it holds that

$$
O\left(\frac{\beta}{\sqrt{\lambda_{t,1}}} \right) \leq \frac{\beta^2}{2\lambda_{t,d}},
\tag{156}
$$

43

Plugging (156) into (154) gives

$$\left(\nu_{t,1} + \frac{\beta w_{t,1}}{\sqrt{\lambda_{t,1}}}\right)^2 + \max_{\|\boldsymbol{w}_{-1}\|_2 = \sqrt{1-w_1^2}} \sum_{i=2}^d \left(\nu_{t,i} + \frac{\beta w_{t,i}}{\sqrt{\lambda_{t,i}}}\right)^2$$

$$\leq 1 + \frac{\beta^2}{2\lambda_{t,d}} + O\left(\frac{\beta^2}{\lambda_{t,d}}\right) \cdot \sqrt{\sum_{i=2}^d w_{t,i}^2}. \tag{157}$$

Comparing the (155) and (157) ensures that there exists a universal constant $c_0 > 0$ such that

$$\sqrt{\sum_{i=2}^d w_{t,i}^2} \geq c_0. \tag{158}$$

Therefore, by a direct application of Lemma 3, for some constant C_1, we define

$$k_t = \{i : \lambda_{t,i} \leq C_1 \lambda_{t,d}\}.$$

In words, this set indexes the eigenvalues that are not "too large" compared to the smallest one. Then there exists a constant C_2 such that

$$\sum_{i \in \mathcal{L}_t} \kappa_{t,i}^2 \geq C_2 \cdot \sum_{i=2}^d \kappa_{t,i}^2. \tag{159}$$

This ensures that a nontrivial fraction of the total variance (as measured by the $\kappa_{t,i}^2$) is concentrated in the set $\mathcal{L}_t$. We now proceed by following a strategy similar to the one employed in Phase 1. Specifically, we enlarge the set slightly and redefine

$$\mathcal{L}_t = \{i : \lambda_{t,i} \leq 2C_1 \lambda_{t,d}\}.$$

Recall from Lemma 5 that the updated eigenvalues $\lambda_{t+1,1}, \ldots, \lambda_{t+1,d}$ are the roots of the secular equation

$$f(\lambda) = 1 + \sum_{i=1}^d \frac{\kappa_{t,i}^2}{\lambda_{t,i} - \lambda}.$$

To study the behavior of the largest eigenvalues (and control their possible growth), we introduce the auxiliary function

$$f_1(\lambda) = 1 + \sum_{i=1}^{k_t} \frac{\kappa_{t,i}^2}{\lambda_{t,i} - \lambda},$$

which only accounts for the "small" eigenvalues indexed by k_t. The idea is that contributions from the larger eigenvalues (outside this set) can be bounded separately. For any eigenvalue $\lambda_{t+1,i}$ with $i \leq j_t$, we have

$$f(\lambda_{t+1,i}) = f_1(\lambda_{t+1,i}) + \sum_{i=k_t+1}^d \frac{\kappa_{t,i}^2}{\lambda_{t,i} - \lambda} \geq f_1(\lambda_{t+1,i}) - \frac{1}{C_1\lambda_{t,d}} \sum_{i=k_t+1}^d \kappa_{t,i}^2. \tag{160}$$

Here, the last inequality comes from the fact that each denominator is at least $C_1\lambda_{t,d}$. Thus, it follows that

$$f_1(\lambda_{t+1,i}) \leq \frac{1}{C_1} \sum_{i=k_t+1}^d \kappa_{t,i}^2. \tag{161}$$

Intuitively, this means that the influence of the "large" eigenvalues (outside k_t) limits how big the roots of f_1 can be. Next, define $\widetilde{\lambda}_{t,1}, \ldots, \widetilde{\lambda}_{t,k_t}$ as the solutions of

$$f_1(\lambda) = \frac{1}{C_1} \sum_{i=k_t+1}^d \kappa_{t,i}^2. \tag{162}$$

44

Equivalently, these values are the roots of the polynomial identity

$$f_1(\lambda) - \frac{1}{C_1} \sum_{i=k_t+1}^{d} \kappa_{t,i}^2 = 1 - \frac{1}{C_1} \sum_{i=k_t+1}^{d} \kappa_{t,i}^2 + \sum_{i=1}^{k_t} \frac{\kappa_{t,i}^2}{\lambda_{t,i} - \lambda}$$

$$= \frac{\left(1 - \frac{1}{C_1} \sum_{i=k_t+1}^{d} \kappa_{t,i}^2\right) \cdot \prod_{i=1}^{k_t}(\lambda_{t,i} - \lambda) + \sum_{i=1}^{k_t} \kappa_{t,i}^2 \prod_{j \neq i}^{k_t}(\lambda_{t,j} - \lambda)}{\prod_{i=1}^{k_t}(\lambda_{t,i} - \lambda)}. \tag{163}$$

By examining the coefficients of this polynomial (via Vieta's formulas), we obtain:

$$m_1 = (-1)^{k_t} \left(1 - \frac{1}{C_1} \sum_{i=k_t+1}^{d} \kappa_{t,i}^2\right),$$

$$m_2 = (-1)^{k_t - 1} \left[\left(1 - \frac{1}{C_1} \sum_{i=k_t+1}^{d} \kappa_{t,i}^2\right) \cdot \sum_{i=1}^{k_t} \lambda_i + \sum_{i=1}^{k_t} \kappa_{t,i}^2\right].$$

This gives the bound

$$\sum_{i=1}^{k_t} \widetilde{\lambda}_{t+1,k_t} = -\frac{m_2}{m_1} = \sum_{i=1}^{k_t} \lambda_i + \frac{\sum_{i=1}^{k_t} \kappa_{t,i}^2}{1 - \frac{1}{C_1} \sum_{i=k_t+1}^{d} \kappa_{t,i}^2} \leq \sum_{i=1}^{k_t} \lambda_i + \frac{1 - C_2 \cdot \sum_{i=2}^{d} \kappa_{t,i}^2}{1 - \frac{C_2}{C_1} \cdot \sum_{i=2}^{d} \kappa_{t,i}^2}. \tag{164}$$

This shows that the sum of these "controlled" eigenvalues cannot grow too quickly, since the denominator penalizes large contributions from $\sum_{i=2}^{d} \kappa_{t,i}^2$. From (161), we know that for all $i \leq j_t$,

$$\lambda_{t+1,i} \leq \widetilde{\lambda}_{t+1,i}.$$

Hence, the sum of the first j_t eigenvalues of $\mathbf{\Lambda}_T$ can be bounded by

$$\sum_{i=1}^{j_t} \lambda_{t+1,i} \leq \sum_{i=1}^{j_t} \widetilde{\lambda}_{t+1,i} = \sum_{i=1}^{j_t} \lambda_{t,i} + \sum_{i=1}^{j_t} (\widetilde{\lambda}_{t+1,i} - \lambda_{t,i})$$

$$\leq \sum_{i=1}^{j_t} \lambda_{t,i} + \frac{1 - C_2}{1 - \frac{C_2}{C_1}} \leq \sum_{i=1}^{j_t} \lambda_{t,i} + \frac{1 - C_2 \sum_{i=2}^{d} \kappa_{t,i}^2}{1 - \frac{C_2}{2} \cdot \sum_{i=2}^{d} \kappa_{t,i}^2}$$

$$\leq \sum_{i=1}^{j_t} \lambda_{t,i} + 1 - \frac{C_2}{2} \cdot \sum_{i=2}^{d} \kappa_{t,i}^2. \tag{165}$$

The key takeaway is that the leading eigenvalues are essentially "capped" in their growth: they can increase by at most a bounded additive amount, while the remaining eigenvalues must absorb a proportional share of the increase. This balance becomes explicit when we look at the complement:

$$\sum_{i=j_t+1}^{d} \lambda_{t+1,i} \geq \sum_{i=j_t+1}^{d} \lambda_{t,i} + \frac{C_2}{2} \cdot \sum_{i=2}^{d} \kappa_{t,i}^2. \tag{166}$$

That is, the "smaller" eigenvalues (those beyond j_t) are guaranteed to grow by a nontrivial amount. Moreover, since

$$\lambda_{t+1,1} \geq \lambda_{t,1} + \kappa_{t,1}^2, \tag{167}$$

we also have

$$\sum_{i=2}^{d} \lambda_{t+1,i} \leq \sum_{i=2}^{d} \lambda_{t,i} + \sum_{i=2}^{d} \kappa_{t,i}^2. \tag{168}$$

Putting everything together, we finally obtain

$$\sum_{i \in \mathcal{L}_t} (\lambda_{t+1,i} - \lambda_{t,i}) \geq \frac{C_2}{2} \cdot \sum_{i=2}^{d} \kappa_{t,i}^2 \geq \frac{C_2}{2} \cdot \sum_{i=2}^{d} (\lambda_{t+1,i} - \lambda_{t,i}). \tag{169}$$

Step 4: lower bound the minimum eigenvalue. We are now in position to establish the main result of this phase. Specifically, we show that when $\underline{c}_t$ is sufficiently small and throughout any period in which $c_t/\underline{c}_t \leq 2$, the minimum eigenvalue $\lambda_{t,d}$ grows at least on the order of $O(\beta\sqrt{t}/\sqrt{d})$. In the previous step we observed that the threshold $\widetilde{c}$ can be chosen arbitrarily small. We therefore set $\widetilde{c} = O(d^{-1/2})$ and define

$$\mathcal{S} = \left\{ t : c_t \leq \widetilde{c}, \ \frac{c_t}{\underline{c}_t} \leq 2 \right\}.$$

By construction, $\mathcal{S}$ is a finite union of *contiguous integer intervals* (segments) of time. Our objective is to obtain a uniform lower bound on c_t (and hence on $\lambda_{t,d}$); it suffices to analyze c_t on each segment in $\mathcal{S}$ separately. Throughout, recall that $\overline{\lambda}_t$ denotes the average of the nonleading eigenvalues (i.e., $\overline{\lambda}_t = \frac{1}{d-1}\sum_{i=2}^{d}\lambda_{t,i}$) and that $C_1, C_2, C_3 > 0$ are absolute constants independent of t and d.

Fix any segment and let $t' < t''$ be its endpoints. For any $t \in [t', t'']$, we compare lower and upper bounds on the cumulative change of the $\lambda_{t,i}$. Summing the growth of all but the smallest eigenvalue yields

$$\sum_{t=t'}^{t''-1} \sum_{i:\lambda_{t,i}\in\mathcal{L}_t} (\lambda_{t+1,i} - \lambda_{t,i}) \geq C_2 \sum_{t=t'}^{t''-1} \sum_{i=2}^{d} (\lambda_{t+1,i} - \lambda_{t,i})$$

$$= C_2 \sum_{i=2}^{d} (\lambda_{t'',i} - \lambda_{t',i})$$

$$= C_2(d-1)(\overline{\lambda}_{t''} - \overline{\lambda}_{t'}). \tag{170}$$

For the matching upper bound, define

$$\widetilde{\mathcal{L}}_{t,t''} = \{i : \lambda_{t,i} \leq C_1\lambda_{t'',d}\},$$

and apply the same counting argument used in Phase #1 (controlling how many coordinates can exceed a multiple of the minimum). This gives

$$\sum_{t=t'}^{t''-1} \sum_{i:\lambda_{t,i}\in\mathcal{L}_t} (\lambda_{t+1,i} - \lambda_{t,i}) \leq \sum_{t=t'}^{t''-1} \sum_{i\in\widetilde{\mathcal{L}}_{t,t''}} (\lambda_{t+1,i} - \lambda_{t,i}) \leq (d-1)\left(C_1\lambda_{t'',d} + 1 - \lambda_{t',d}\right). \tag{171}$$

Combining (170) and (171) yields

$$C_2(\overline{\lambda}_{t''} - \overline{\lambda}_{t'}) \leq C_1\lambda_{t'',d} + 1 - \lambda_{t',d},$$

and hence

$$\lambda_{t'',d} \geq \min\left(\lambda_{t',d}, \ \frac{C_2}{C_1}(\overline{\lambda}_{t''} - \overline{\lambda}_{t'}) + \frac{\lambda_{t',d} - 1}{C_1}\right). \tag{172}$$

We next control the minimum of c_t on each segment of $\mathcal{S}$. Without loss of generality assume $t_1 \notin \mathcal{S}$ (this can be arranged by taking $\widetilde{c}$ below a fixed absolute constant), so every segment begins at some $t' > t_0$. There are two ways a new segment can start:

(i) $c_{t'} \leq \widetilde{c}$ while $c_{t'-1} > \widetilde{c}$. By the $O(t^{-1/2})$ drift established earlier for c_t (smooth variation), the threshold crossing cannot overshoot by more than a fixed fraction for large t', hence $c_{t'} \geq 0.95\widetilde{c}$.

(ii) $c_{t'}/\underline{c}_{t'} \leq 2$ while $c_{t'-1}/\underline{c}_{t'-1} > 2$. The same smoothness argument applied to the ratio shows $c_{t'}/\underline{c}_{t'} > 1.9$, so $c_{t'} \geq 1.9\underline{c}_{t'}$ for large t'.

Combining (i)–(ii) gives the useful entry condition

$$c_{t'} \geq \min\left(1.9\underline{c}_{t'}, \ 0.95\widetilde{c}\right).$$

From the growth bounds of the preceding steps, for all t on the same segment as t' we have

$$\overline{\lambda}_T \geq \overline{\lambda}_t + \frac{C_3}{d-1} \cdot \frac{\beta}{2\underline{c}_{t'}\sqrt{t}} \geq \overline{\lambda}_t + \frac{C_3}{d-1} \cdot \frac{\beta}{2\widetilde{c}\sqrt{t}},$$

and summing from t' to t'' yields

$$\overline{\lambda}_{t''} - \overline{\lambda}_{t'} \geq \frac{C_3\beta}{\widetilde{c}(d-1)}(\sqrt{t''} - \sqrt{t'}).$$

Plugging this into (172) and then applying it with $t'' = t$ (for any $t \in [t', t'']$) gives

$$
\begin{aligned}
\lambda_{t,d} &\geq \max\left(\lambda_{t',d}, \ \frac{C_2 C_3 \beta}{C_1 \widetilde{c}(d-1)}(\sqrt{t} - \sqrt{t'}) + \frac{c_{t'}\beta\sqrt{t'} - 1}{C_1}\right) \\
&= \max\left(\lambda_{t',d}, \ c\beta(\sqrt{t} - \sqrt{t'}) + \frac{\min(1.9\underline{c}_{t'}, 0.95\widetilde{c})\beta\sqrt{t'} - 1}{C_1}\right),
\end{aligned}
\tag{173}
$$

where we set

$$c := \frac{C_2 C_3}{C_1 \widetilde{c}(d-1)} = O(d^{-1/2}).$$

We now choose $c^\star = O(d^{-1/2})$ with $c^\star \leq 0.5\,\widetilde{c}$ such that, for all $t \geq t'$,

$$c^\star\beta\sqrt{t} \leq \max\left(1.9c^\star\beta\sqrt{t'}, \ c\beta(\sqrt{t} - \sqrt{t'}) + \frac{1.9c^\star\beta\sqrt{t'} - 1}{C_1}\right).$$

Such a choice is always possible: the right-hand side is the maximum of two affine functions of $\sqrt{t}$ whose slopes are 0 and $c\beta > 0$, respectively, whereas the left-hand side has slope $c^\star\beta$; taking $c^\star \leq c$ and adjusting the intercept via the $-(1/C_1)$ term ensures the inequality holds for all $t \geq t'$. Therefore, for all $t > t'$ on the same segment in $\mathcal{S}$,

$$
\begin{aligned}
\lambda_{t,d} &\geq \max\left(\min(1.9\underline{c}_{t'}, 0.95\widetilde{c})\beta\sqrt{t'}, \ c\beta(\sqrt{t} - \sqrt{t'}) + \frac{\min(1.9\underline{c}_{t'}, 0.95\widetilde{c})\beta\sqrt{t'} - 1}{C_1}\right) \\
&\geq \max\left(1.9c^\star\beta\sqrt{t'}, \ c\beta(\sqrt{t} - \sqrt{t'}) + \frac{1.9c^\star\beta\sqrt{t'} - 1}{C_1}\right) \\
&\geq c^\star\beta\sqrt{t}.
\end{aligned}
\tag{174}
$$

Applying this argument on every segment of $\mathcal{S}$, we conclude that

$$\lambda_{t,d} \geq c^\star\beta\sqrt{t} \asymp \frac{\beta\sqrt{t}}{\sqrt{d}}, \quad \text{for all } t \geq t_1.$$

Finally, from equations (135) and (139),

$$\|\boldsymbol{v}_{t,1} - \boldsymbol{\theta}^\star\|_2^2 = O\left(\frac{\beta\sqrt{d}}{\sqrt{t}}\right), \quad \text{and} \quad \overline{\lambda}_t = O\left(\frac{\beta\sqrt{t}}{\sqrt{d}}\right).$$

Combining these with the lower bound on $\lambda_{t,d}$ established above gives

$$\lambda_{t,d} \asymp \frac{\beta\sqrt{t}}{\sqrt{d}}, \tag{175}$$

and hence the estimation error of the leading eigenvector satisfies

$$\|\boldsymbol{v}_{t,1} - \boldsymbol{\theta}^\star\|_2 = O\left(\frac{\beta}{\sqrt{\lambda_{t,d}}}\right). \tag{176}$$

D.4 Analysis of Phase #3 (proof of Proposition 3)

Step 1: characterizing the update of the next top eigenvector $v_{t+1,1}$. We aim to show in this phase that $v_{t+1,1}$ can be characterized by the previous leading eigenvector $v_{t,1}$ and the estimated parameter $\widehat{\theta}_t$ as follows

$$v_{t+1,1} \;=\; v_{t,1} \;+\; \frac{\widehat{\theta}_t - v_{t,1}}{t} \;+\; \zeta_t,$$

where ζ_t is a high-order perturbation that can be nicely bounded both in norm and direction, being near orthogonal to $v_{t,1} - \widehat{\theta}_t$. We proceed with the following steps to show this result.

Lower bound on $|\kappa_{t,i}|$ $(i \geq 2)$. To make this precise, we first derive a uniform upper bound on

$$\left\| \widehat{\theta}_t + \beta \Lambda_t^{-1/2} w \right\|_2,$$

valid for any unit vector w, i.e., $\|w\|_2 = 1$. We first establish the uniform bound. Expanding the squared norm gives

$$\left\| \widehat{\theta}_t + \beta \Lambda_t^{-1/2} w \right\|_2^2 \;=\; \sum_{i=1}^{d} \left(\nu_{t,i} + \frac{\beta w_i}{\sqrt{\lambda_{t,i}}} \right)^2, \tag{177}$$

where $\nu_{t,i}$ denotes the i-th coordinate of $\widehat{\theta}_t$ expressed in the eigenbasis of Λ_t. For the *leading coordinate* $(i = 1)$, our earlier estimates imply

$$\nu_{t,1} + \frac{\beta w_1}{\sqrt{\lambda_{t,1}}} \;=\; 1 - O\left(\frac{\beta^2}{\lambda_{t,d}}\right) + O\left(\frac{\beta}{\sqrt{t}}\right) \;=\; 1 + O\left(\frac{\beta^2}{\lambda_{t,d}}\right),$$

where we have used the bound $\lambda_{t,d} \lesssim \beta\sqrt{t}/\sqrt{d}$ in Proposition 2 to absorb the $O(\beta/\sqrt{t})$ term into the $O(\beta^2/\lambda_{t,d})$ error. This shows that the leading component remains close to 1, with only a small perturbation.

For the *remaining coordinates* $(i \geq 2)$, we control their contribution by

$$\sum_{i=2}^{d} \left(\nu_{t,i} + \frac{\beta w_i}{\sqrt{\lambda_{t,i}}} \right)^2 \;\lesssim\; \sum_{i=2}^{d} \left(\nu_{t,i}^2 + \frac{\beta^2 w_i^2}{\lambda_{t,i}} \right) \;\leq\; (1 - \nu_{t,1}^2) + \frac{\beta^2}{\lambda_{t,d}} \sum_{i=2}^{d} w_i^2$$

$$= O\left(\frac{\beta^2}{\lambda_{t,d}}\right),$$

where we used $\sum_{i=1}^{d} w_i^2 = 1$ and the fact that $\nu_{t,1}^2 = 1 - O(\beta^2/\lambda_{t,d})$. Putting the two pieces together, we obtain the *uniform expansion*

$$\sum_{i=1}^{d} \left(\nu_{t,i} + \frac{\beta w_i}{\sqrt{\lambda_{t,i}}} \right)^2 \;=\; 1 + O\left(\frac{\beta^2}{\lambda_{t,d}}\right), \tag{178}$$

valid uniformly over all unit vectors w and all $t \geq t_1$. We now use (178) to sharpen the characterization of the maximizer a_t. Recall its definition:

$$w_t \;=\; \arg \max_{\|w\|_2=1} \sum_{i=1}^{d} \left(\nu_{t,i} + \frac{\beta w_i}{\sqrt{\lambda_{t,i}}} \right)^2 \;=\; \arg \max_{\|w\|_2=1} \sum_{i=1}^{d} \left(\frac{2\beta \nu_{t,i} w_{t,i}}{\sqrt{\lambda_{t,i}}} + \frac{\beta^2 w_{t,i}^2}{\lambda_{t,i}} \right).$$

The maximization is dominated by the cross-term

$$2 \sum_{i=1}^{d} \frac{\beta}{\sqrt{\lambda_{t,i}}} \nu_{t,i} w_i,$$

which is maximized when $w_{t,i}$ aligns in sign with $\nu_{t,i}$. Hence, at the maximizer we necessarily have

$$w_{t,i} \nu_{t,i} \;\geq\; 0 \quad \text{for all } i.$$

By (178), we know that for any unit vector $\boldsymbol{w}$—in particular, for $\boldsymbol{w} = \boldsymbol{w}_t$—the perturbation vector satisfies

$$\left\|\widehat{\boldsymbol{\theta}}_t + \beta \boldsymbol{\Lambda}_t^{-1/2} \boldsymbol{w}\right\|_2 = 1 + O\left(\frac{\beta^2}{\lambda_{t,d}}\right).$$

Consequently, the normalized projection operator satisfies

$$\mathcal{P}\left(\widehat{\boldsymbol{\theta}}_t + \beta \boldsymbol{\Lambda}_t^{-1/2} \boldsymbol{w}\right) = \left(1 - O\left(\frac{\beta^2}{\lambda_{t,d}}\right)\right) \cdot \left(\widehat{\boldsymbol{\theta}}_t + \beta \boldsymbol{\Lambda}_t^{-1/2} \boldsymbol{w}\right). \tag{179}$$

For each coordinate $i \in [d]$, this implies the refined characterization

$$\kappa_{t,i} = \left(1 - O\left(\frac{\beta^2}{\lambda_{t,d}}\right)\right) \left(\nu_{t,i} + \frac{\beta w_{t,i}}{\sqrt{\lambda_{t,i}}}\right). \tag{180}$$

Since $w_{t,i}\nu_{t,i} \geq 0$, the correction preserves the sign of the signal and enlarges its magnitude:

$$\left|\nu_{t,i} + \frac{\beta w_{t,i}}{\sqrt{\lambda_{t,i}}}\right| \geq |\nu_{t,i}|.$$

Therefore,

$$|\kappa_{t,i}| = \left(1 - O\left(\frac{\beta^2}{\lambda_{t,d}}\right)\right) \cdot \left|\nu_{t,i} + \frac{\beta w_{t,i}}{\sqrt{\lambda_{t,i}}}\right| \geq \left(1 - O\left(\frac{\beta^2}{\lambda_{t,d}}\right)\right) |\nu_{t,i}|. \tag{181}$$

In other words, the normalized contributions $\kappa_{t,i}$ not only preserve the alignment with the underlying signal but also lose at most an $O(\beta^2/\lambda_{t,d})$ fraction of their magnitude. This guarantees stability of the signal direction under the perturbation.

Spectral decomposition of next top eigenvector $\boldsymbol{v}_{t+1,1}$. We now decompose the next top eigenvector $\boldsymbol{v}_{t+1,1}$ in the eigenbasis formed by the previous eigenvectors $\{\boldsymbol{v}_{t,1}, \ldots, \boldsymbol{v}_{t,d}\}$. That is, we write

$$\boldsymbol{v}_{t+1,1} = \sum_{i=1}^{d} \omega_{t,i} \boldsymbol{v}_{t,i}. \tag{182}$$

Our goal is to characterize the coefficients $\{\omega_{t,i}\}$, in particular establishing a nontrivial lower bound for $\omega_{t,i}$ when $i \geq 2$. From Lemma 5, the top eigenvector after the rank-one perturbation can be expressed explicitly as

$$\boldsymbol{v}_{t+1,1} = K_t^{-1} \sum_{i=1}^{d} \frac{\kappa_{t,i}}{\lambda_{t+1,1} - \lambda_{t,i}} \boldsymbol{v}_{t,i}, \tag{183}$$

where K_t is a normalization constant ensuring $\|\boldsymbol{v}_{t+1,1}\|_2 = 1$. In particular,

$$K_t = \left(\sum_{i=1}^{d} \left(\frac{\kappa_{t,i}}{\lambda_{t+1,1} - \lambda_{t,i}}\right)^2\right)^{1/2}.$$

Intuitively, the denominator $\lambda_{t+1,1} - \lambda_{t,i}$ captures the spectral separation between the new leading eigenvalue $\lambda_{t+1,1}$ and the old eigenvalues $\{\lambda_{t,i}\}$, while the numerator $\kappa_{t,i}$ measures the alignment between the perturbation direction and $\boldsymbol{v}_{t,i}$. Thus, the size of each $\omega_{t,i}$ is governed both by spectral gaps and by how much the perturbation projects onto $\boldsymbol{v}_{t,i}$.

We first control the normalization factor K_t. From Lemma 6, the increment of the leading eigenvalue satisfies

$$\lambda_{t+1,1} - \lambda_{t,1} = 1 + O\left(\frac{\beta^2}{\lambda_{t,d}}\right).$$

Moreover, for $i \geq 2$, the spectral gap is lower bounded as

$$\lambda_{t+1,1} - \lambda_{t,i} \geq \lambda_{t,1} - \lambda_{t,i} \gtrsim t,$$

thanks to the assumed eigengap structure at time t. Substituting the bounds on $\kappa_{t,i}$ from Lemma 2, we obtain

$$\begin{aligned}
K_t^2 &= \left(\frac{\kappa_{t,1}}{\lambda_{t+1,1} - \lambda_{t,1}} \right)^2 + \sum_{i=2}^{d} \left(\frac{\kappa_{t,i}}{\lambda_{t+1,1} - \lambda_{t,i}} \right)^2 \\
&= \left(1 - O\left(\frac{\beta^2}{\lambda_{t,d}} \right) \right)^2 + O(t^{-2}) \cdot O\left(\frac{\beta^2}{\lambda_{t,d}} \right) \\
&= 1 - O\left(\frac{\beta^2}{\lambda_{t,d}} \right),
\end{aligned} \tag{184}$$

which implies $K_t = 1 - O(\beta^2 / \lambda_{t,d})$. Now extracting coefficients from (183), we see that for $i \geq 2$,

$$\begin{aligned}
|\omega_{t,i}| = K_t^{-1} \cdot \frac{|\kappa_{t,i}|}{\lambda_{t+1,1} - \lambda_{t,i}} \\
\geq \left(1 + O\left(\frac{\beta^2}{\lambda_{t,d}} \right) \right) \cdot \frac{\left(1 - O\left(\frac{\beta^2}{\lambda_{t,d}} \right) \right) |\nu_{t,i}|}{t} \\
\geq \left(1 - O\left(\frac{\beta^2}{\lambda_{t,d}} \right) \right) \cdot \frac{|\nu_{t,i}|}{t}.
\end{aligned} \tag{185}$$

Thus, a clean lower bound is established for all $\omega_{t,i}$ with $i \geq 2$. Since $\kappa_{t,i} \nu_{t,i} \geq 0$ and $\omega_{t,i}$ inherits the sign of $\kappa_{t,i}$, we further deduce that

$$\omega_{t,i} \nu_{t,i} \geq 0, \qquad \text{for all } i \geq 2. \tag{186}$$

Turning to the leading coefficient $\omega_{t,1}$, observe that

$$\sum_{i=2}^{d} \omega_{t,i}^2 = K_t^{-2} \cdot \sum_{i=2}^{d} \left(\frac{\kappa_{t,i}}{\lambda_{t+1,1} - \lambda_{t,i}} \right)^2 \lesssim \left(1 + O\left(\frac{\beta^2}{\lambda_{t,d}} \right) \right) \cdot \frac{O(\beta^2 / \lambda_{t,d})}{t^2} = O\left(\frac{\beta^2}{t^2 \lambda_{t,d}} \right). \tag{187}$$

Hence,

$$\omega_{t,1} = \sqrt{1 - \sum_{i=2}^{d} \omega_{t,i}^2} = 1 - O\left(\frac{\beta^2}{t^2 \lambda_{t,d}} \right). \tag{188}$$

Characterize the norm and direction of ζ_t. To obtain a recursive characterization of $v_{t+1,1}$, it remains to compare it against the "linearized" update in the direction of $\widehat{\theta}_t$. Specifically, define the intermediate vector

$$\overline{v}_{t+1,1} := v_{t,1} + \frac{\widehat{\theta}_t - v_{t,1}}{t}, \tag{189}$$

which lies on the line segment connecting $v_{t,1}$ and $\widehat{\theta}_t$. Expanding this vector in the eigenbasis $\{v_{t,1}, \ldots, v_{t,d}\}$ gives

$$\overline{v}_{t+1,1} = \left(1 - \frac{1 - \nu_{t,1}}{t} \right) v_{t,1} + \sum_{i=2}^{d} \frac{\nu_{t,i}}{t} v_{t,i}, \tag{190}$$

where $\widehat{\theta}_t = \sum_{i=1}^{d} \nu_{t,i} v_{t,i}$ is the eigenbasis decomposition of $\widehat{\theta}_t$.

We now measure the discrepancy between $\boldsymbol{v}_{t+1,1}$ and this linearized vector:

$$\|\boldsymbol{v}_{t+1,1} - \overline{\boldsymbol{v}}_{t+1,1}\|_2^2 = \left(\omega_{t,1} - 1 + \frac{1 - \nu_{t,1}}{t}\right)^2 + \sum_{i=2}^{d} \left(\omega_{t,i} - \frac{\nu_{t,i}}{t}\right)^2.$$

Using the bounds established earlier on $\omega_{t,1}$ and $\omega_{t,i}$, we obtain

$$\begin{aligned}
\|\boldsymbol{v}_{t+1,1} - \overline{\boldsymbol{v}}_{t+1,1}\|_2^2 &\le \left(O\left(\frac{\beta^2}{t^2 \lambda_{t,d}}\right) + O\left(\frac{\beta^2}{t \lambda_{t,d}}\right)\right)^2 + \sum_{i=2}^{d} \omega_{t,i}^2 \\
&\quad + \sum_{i=2}^{d} \left(\frac{\nu_{t,i}}{t} - \left(1 - O\left(\frac{\beta^2}{\lambda_{t,d}}\right)\right) \cdot \frac{\nu_{t,i}}{t}\right)^2 \\
&\le O\left(\frac{\beta^4}{t^2 \lambda_{t,d}^2}\right) + O\left(\frac{\beta^4}{t^2 \lambda_{t,d}^2}\right) \\
&= O\left(\frac{\beta^4}{t^2 \lambda_{t,d}^2}\right). \tag{191}
\end{aligned}$$

This uses the inequality

$$\left|\omega_{t,i} - \frac{\nu_{t,i}}{t}\right| \le \max\left\{|\omega_{t,i}|, \left|\frac{\nu_{t,i}}{t} - \left(1 - O\left(\frac{\beta^2}{\lambda_{t,d}}\right)\right) \cdot \frac{\nu_{t,i}}{t}\right|\right\},$$

which in turn implies

$$\sum_{i=2}^{d} \left(\omega_{t,i} - \frac{\nu_{t,i}}{t}\right)^2 \le \sum_{i=2}^{d} \omega_{t,i}^2 + \sum_{i=2}^{d} \left(\frac{\nu_{t,i}}{t} - \left(1 - O\left(\frac{\beta^2}{\lambda_{t,d}}\right)\right) \cdot \frac{\nu_{t,i}}{t}\right)^2.$$

Therefore, we have shown that

$$\|\boldsymbol{\zeta}_t\|_2 = \|\boldsymbol{v}_{t+1,1} - \overline{\boldsymbol{v}}_{t+1,1}\|_2 = O\left(\frac{\beta^2}{t \lambda_{t,d}}\right). \tag{192}$$

Next, we analyze the correlation between the perturbation $\boldsymbol{\zeta}_t$ and the deviation vector $\boldsymbol{v}_{t,1} - \widehat{\boldsymbol{\theta}}_t$. Recalling that

$$\widehat{\boldsymbol{\theta}}_t = \sum_{i=1}^{d} \nu_{t,i} \boldsymbol{v}_{t,i},$$

we compute

$$\left\langle \boldsymbol{\zeta}_t, \boldsymbol{v}_{t,1} - \widehat{\boldsymbol{\theta}}_t \right\rangle = (1 - \nu_{t,1}) \cdot \left(\omega_{t,1} - 1 + \frac{1 - \nu_{t,1}}{t}\right) - \sum_{i=2}^{d} \left(\omega_{t,i} - \frac{\nu_{t,i}}{t}\right) \nu_{t,i}. \tag{193}$$

From (185), we know that for $i \ge 2$,

$$|\omega_{t,i}| \ge \left(1 - O\left(\frac{\beta^2}{\lambda_{t,d}}\right)\right) \cdot \frac{|\nu_{t,i}|}{t}, \quad \text{and} \quad \omega_{t,i} \nu_{t,i} \ge 0.$$

This leads to the bound

$$\left(\omega_{t,i} - \frac{\nu_{t,i}}{t}\right) \nu_{t,i} \ge O\left(\frac{\beta^2}{t \lambda_{t,d}}\right) \cdot \nu_{t,i}^2.$$

Consequently, one can conclude

$$\left\langle \boldsymbol{\zeta}_t, \boldsymbol{v}_{t,1} - \widehat{\boldsymbol{\theta}}_t \right\rangle \le O\left(\frac{\beta^2}{\lambda_{t,d}}\right) \cdot O\left(\frac{\beta^2}{t \lambda_{t,d}} + \frac{\beta^2}{t^2 \lambda_{t,d}}\right) + O\left(\frac{\beta^2}{t \lambda_{t,d}}\right) \cdot \sum_{i=2}^{d} \nu_{t,i}^2$$

51

$$= O\left(\frac{\beta^4}{t\lambda_{t,d}^2}\right). \tag{194}$$

To summarize, we have obtained the desired recursive form:

$$\boldsymbol{v}_{t+1,1} = \boldsymbol{v}_{t,1} + \frac{\widehat{\boldsymbol{\theta}}_t - \boldsymbol{v}_{t,1}}{t} + \boldsymbol{\zeta}_t, \tag{195}$$

where the perturbation satisfies

$$\|\boldsymbol{\zeta}_t\|_2 = O\left(\frac{\beta^2}{t\lambda_{t,d}}\right), \qquad \left\langle \boldsymbol{\zeta}_t, \boldsymbol{v}_{t,1} - \widehat{\boldsymbol{\theta}}_t \right\rangle = O\left(\frac{\beta^4}{t\lambda_{t,d}^2}\right).$$

This fine-grained control of $\boldsymbol{\zeta}_t$ provides a precise recursive characterization of the leading eigenvector and underpins the convergence analysis of $\boldsymbol{v}_{t,1}$.

Step 2: establishing the inductive concentration of the leading eigenvector. In this final step, we complete the concentration analysis by showing that the leading eigenvector $\boldsymbol{v}_{t,1}$ converges to the true signal direction $\boldsymbol{\theta}^\star$ at the desired rate. Our goal is to control the error term $\left\|\boldsymbol{v}_{t,1} - \boldsymbol{\theta}^\star\right\|_2$ for sufficiently large t. Building on the update characterization derived in the previous step, we relate the deviation at time $t+1$ to that at time t. In particular, we obtain the recursive inequality

$$\left\|\boldsymbol{v}_{t+1,1} - \boldsymbol{\theta}^\star\right\|_2^2 \leq \left(\left(1 - \frac{1}{t}\right)\left\|\boldsymbol{v}_{t,1} - \boldsymbol{\theta}^\star\right\|_2 + \widetilde{O}(t^{-5/4})\right)^2 + \widetilde{O}(t^{-2}),$$

which captures the contraction of the eigenvector error up to higher-order perturbation terms. This recurrence illustrates a decaying trend: the leading error shrinks multiplicatively by approximately $(1 - 1/t)$ at each iteration, while additive fluctuations vanish at a faster polynomial rate. By applying a careful induction argument and leveraging the initialization guarantee established in Phase #2, we deduce that there exists t_2 such that for all $t \geq t_2$, the top eigenvector achieves the desired concentration, thereby converging toward the ground-truth direction $\boldsymbol{\theta}^\star$.

To formalize this argument, recall the auxiliary update

$$\overline{\boldsymbol{v}}_{t+1,1} := \boldsymbol{v}_{t,1} + \frac{\widehat{\boldsymbol{\theta}}_t - \boldsymbol{v}_{t,1}}{t},$$

which enables a convenient decomposition of the error:

$$\begin{aligned}
\overline{\boldsymbol{v}}_{t+1,1} - \boldsymbol{\theta}^\star &= (\overline{\boldsymbol{v}}_{t+1,1} - \widehat{\boldsymbol{\theta}}_t) + (\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^\star) \\
&= \left(1 - \frac{1}{t}\right) \cdot (\boldsymbol{v}_{t,1} - \boldsymbol{\theta}^\star) + \frac{1}{t} \cdot (\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^\star).
\end{aligned} \tag{196}$$

From this relation, we immediately obtain the inequality

$$\|\overline{\boldsymbol{v}}_{t+1,1} - \boldsymbol{\theta}^\star\|_2 \leq \left(1 - \frac{1}{t}\right)\|\boldsymbol{v}_{t,1} - \boldsymbol{\theta}^\star\|_2 + \frac{c}{t}\sqrt{\frac{d + \log\log T}{\lambda_{t,d}}}. \tag{197}$$

Next, analyzing the correction term $\boldsymbol{\zeta}_t$ yields

$$\begin{aligned}
\|\boldsymbol{v}_{t+1,1} - \boldsymbol{\theta}^\star\|_2^2 &= \|\overline{\boldsymbol{v}}_{t+1,1} - \boldsymbol{\theta}^\star\|_2^2 + 2\langle\overline{\boldsymbol{v}}_{t+1,1} - \boldsymbol{\theta}^\star, \boldsymbol{\zeta}_t\rangle + \|\boldsymbol{\zeta}_t\|_2^2 \\
&\leq \left(\left(1 - \frac{1}{t}\right)\|\boldsymbol{v}_{t,1} - \boldsymbol{\theta}^\star\|_2 + \frac{c\sigma}{t}\sqrt{\frac{d + \log\log T}{\lambda_{t,d}}}\right)^2 + O\left(\frac{\beta^4}{t\lambda_{t,d}^2}\right) \\
&= \left(\left(1 - \frac{1}{t}\right)\|\boldsymbol{v}_{t,1} - \boldsymbol{\theta}^\star\|_2 + \frac{1}{t^{5/4}} \cdot \frac{cd^{1/4}(\sigma\sqrt{d + \log\log T} + 1)}{\sqrt{\beta}}\right)^2 + O\left(\frac{\beta^2 d}{t^2}\right),
\end{aligned} \tag{198}$$

where the last equality follows since $\lambda_{t,d} \asymp \beta\sqrt{t}/\sqrt{d}$.

To bound this sequence rigorously, we invoke the following key lemma.

Lemma 10. *Let $\{a_n\}$ be a sequence satisfying*

$$a_{n+1}^2 \leq \left(\left(1 - \frac{1}{n}\right) a_n + \frac{B}{n^{5/4}} \right)^2 + \frac{C}{n^2},$$

for constants B, C. Fix $n_0 \geq \max\left(16, 9C^2/(16B^4)\right)$. Then, for any $n \geq (a_{n_0}^2 n_0^{3/2})/(16B^2)$,

$$a_n \leq \frac{4B}{n^{1/4}}. \tag{199}$$

The proof of Lemma 10 is deferred to Appendix E. Applying this lemma with

$$B = \frac{cd^{1/4}(\sigma\sqrt{d + \log\log T} + 1)}{\sqrt{\beta}}, \qquad C = O(\beta^2 d),$$

we obtain that, setting

$$t_1' = O\left(\frac{C^2}{B^4}\right) = O\left(\frac{\beta^6}{\sigma^4 d}\right),$$

and defining

$$t_2 = \frac{\|\boldsymbol{v}_{t_1',1} - \boldsymbol{\theta}^\star\|_2^2 \cdot t_1'^{3/2}}{16B^2} = O\left(\frac{\beta^8}{\sigma^6 d^2}\right), \tag{200}$$

we have for all $t \geq t_2$ that

$$\|\boldsymbol{v}_{t,1} - \boldsymbol{\theta}^\star\|_2 \lesssim \frac{B}{t^{1/4}} = O\left(\frac{\sigma\sqrt{d + \log\log T} + 1}{\sqrt{\lambda_{t,d}}}\right).$$

Finally, combining this with the deviation of $\widehat{\boldsymbol{\theta}}_t$, we obtain

$$\|\boldsymbol{v}_{t+1,1} - \widehat{\boldsymbol{\theta}}_t\|_2 \leq \|\boldsymbol{v}_{t+1,1} - \boldsymbol{\theta}^\star\|_2 + \|\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^\star\|_2 = O\left(\frac{\sigma\sqrt{d + \log\log T} + 1}{\sqrt{\lambda_{t,d}}}\right),$$

which establishes the final concentration bound and completes the proof of Proposition 3.

D.5 Analysis of Phase #4 (proof of Proposition 4)

Step 1: precise decomposition of the action vector. Write all vectors in the orthonormal eigenbasis $\{\boldsymbol{v}_{t,1}, \ldots, \boldsymbol{v}_{t,d}\}$ so that $\nu_{t,i}$ are the coordinates of $\boldsymbol{a}_t$ (and similarly for $\widehat{\boldsymbol{\theta}}_t$), $\lambda_{t,i}$ are the eigenvalues, and w_i the coordinates of any unit vector $\boldsymbol{w}$. The objective

$$g_t(\boldsymbol{w}) = \sum_{i=1}^d \left(\nu_{t,i} + \frac{\beta w_i}{\sqrt{\lambda_{t,i}}}\right)^2$$

The maximizer of $g_t(\boldsymbol{w})$, denoted as $\boldsymbol{w}_t$, characterizes the direction of $\boldsymbol{a}_t$. Directly optimizing $g_t(\boldsymbol{w})$ can be difficult in general, so we introduce an approximate objective. We first expand the square, which gives

$$g_t(\boldsymbol{w}) = \sum_{i=1}^d \Big(\underbrace{\nu_{t,i}^2}_{\text{constant in } \boldsymbol{w}} + \underbrace{\frac{2\beta\nu_{t,i}w_i}{\sqrt{\lambda_{t,i}}}}_{\text{linear in } w_i} + \underbrace{\frac{\beta^2 w_i^2}{\lambda_{t,i}}}_{\text{quadratic in } w_i} \Big).$$

Since $\nu_{t,1} \approx 1$ and $\nu_{t,i}$ for $i \geq 2$ are small, we linearize the $i = 1$ term around $\nu_{t,1} = 1$ and drop the (typically smaller) quadratic piece in w_1, while for $i \geq 2$ we drop the tiny linear terms and keep only the quadratic regularization. This yields the tractable surrogate

$$\widetilde{g}_t(\boldsymbol{w}) = 1 + \underbrace{\frac{2\beta w_1}{\sqrt{\lambda_{t,1}}}}_{\text{dominant linear response along } i=1} + \underbrace{\sum_{i=2}^d \frac{\beta^2 w_i^2}{\lambda_{t,i}}}_{\text{penalizes transverse energy}}. \tag{201}$$

Define

$$g_{t,1}(\boldsymbol{w}) := \left(\nu_{t,1} + \frac{\beta w_1}{\sqrt{\lambda_{t,1}}}\right)^2, \qquad\qquad g_{t,2}(\boldsymbol{w}) := \sum_{i=2}^{d} \left(\nu_{t,i} + \frac{\beta w_i}{\sqrt{\lambda_{t,i}}}\right)^2,$$

$$\widetilde{g}_{t,1}(\boldsymbol{w}) := 1 + \frac{2\beta w_1}{\sqrt{\lambda_{t,1}}}, \qquad\qquad \widetilde{g}_{t,2}(\boldsymbol{w}) := \sum_{i=2}^{d} \frac{\beta^2 w_i^2}{\lambda_{t,i}}.$$

By Lemma 2, for

$$h_t = O\left(\frac{\sigma\sqrt{d + \log\log T} + 1}{\sqrt{\lambda_{t,d}}}\right)$$

we have the accuracy guarantees

$$\nu_{t,1} \geq 1 - O\left(\frac{\sigma^2(d + \log\log T) + 1}{\lambda_{t,d}}\right), \qquad \nu_{t,i} = O\left(\frac{\sigma\sqrt{d + \log\log T} + 1}{\sqrt{\lambda_{t,d}}}\right) \quad (i \geq 2).$$

Expanding and regrouping,

$$\left|g_{t,1}(\boldsymbol{w}) - \widetilde{g}_{t,1}(\boldsymbol{w})\right| = \left|\nu_{t,1}^2 + \frac{2\beta w_1 \nu_{t,1}}{\sqrt{\lambda_{t,1}}} + \frac{\beta^2 w_1^2}{\lambda_{t,1}} - 1 - \frac{2\beta w_1}{\sqrt{\lambda_{t,1}}}\right|$$

$$\leq 1 - \nu_{t,1}^2 + (1 - \nu_{t,1})\left|\frac{2\beta w_1}{\sqrt{\lambda_{t,1}}}\right| + \left|\frac{\beta^2 w_1^2}{\lambda_{t,1}}\right|$$

$$= O\left(\frac{\sigma^2(d + \log\log T) + 1}{\lambda_{t,d}}\right)\left(1 + O\left(\frac{\beta}{\sqrt{t}}\right)\right) + O\left(\frac{\beta^2}{t}\right)$$

$$= O\left(\frac{\sigma^2(d + \log\log T) + 1}{\lambda_{t,d}}\right) + O\left(\frac{\beta^2}{t}\right). \tag{202}$$

Here, the first term uses $|1 - \nu_{t,1}^2| = (1 - \nu_{t,1})(1 + \nu_{t,1})$ with $1 - \nu_{t,1} = O(\cdot)$ and $\nu_{t,1} \leq 1$; the second uses $|w_1| \leq 1$; the third simply bounds the quadratic remainder.

For the bound of $|g_{t,2}(\boldsymbol{w}) - \widetilde{g}_{t,2}(\boldsymbol{w})|$, using $(a + b)^2 - b^2 = a(a + 2b)$ with $a = \nu_{t,i}$ and $b = \beta w_i/\sqrt{\lambda_{t,i}}$,

$$\left|g_{t,2}(\boldsymbol{w}) - \widetilde{g}_{t,2}(\boldsymbol{w})\right| = \sum_{i=2}^{d} |\nu_{t,i}| \left|\nu_{t,i} + \frac{2\beta w_i}{\sqrt{\lambda_{t,i}}}\right|$$

$$\leq \sum_{i=2}^{d} \nu_{t,i}^2 + 2\beta \sum_{i=2}^{d} \frac{|\nu_{t,i}||w_i|}{\sqrt{\lambda_{t,i}}} \leq \sum_{i=2}^{d} \nu_{t,i}^2 + 2\beta\sqrt{\sum_{i=2}^{d} \frac{\nu_{t,i}^2}{\lambda_{t,i}}} \sqrt{\sum_{i=2}^{d} w_i^2}.$$

Here, the first inequality is triangle inequality; the second is Cauchy–Schwarz. Since $\|\boldsymbol{w}\|_2 = 1$ and $\lambda_{t,i} \geq \lambda_{t,d}$,

$$\sum_{i=2}^{d} \nu_{t,i}^2 = O\left(\frac{\sigma^2(d + \log\log T) + 1}{\lambda_{t,d}}\right), \qquad \sqrt{\sum_{i=2}^{d} \frac{\nu_{t,i}^2}{\lambda_{t,i}}} \leq \frac{1}{\sqrt{\lambda_{t,d}}}\sqrt{\sum_{i=2}^{d} \nu_{t,i}^2} = O\left(\frac{\sigma\sqrt{d + \log\log T} + 1}{\lambda_{t,d}}\right),$$

which yields

$$\left|g_{t,2}(\boldsymbol{w}) - \widetilde{g}_{t,2}(\boldsymbol{w})\right| = O\left(\frac{\sigma^2(d + \log\log T) + 1}{\lambda_{t,d}}\right) + O\left(\frac{\beta(\sigma\sqrt{d + \log\log T} + 1)}{\lambda_{t,d}}\right)$$

$$= O\left(\frac{\beta(\sigma\sqrt{d + \log\log T} + 1)}{\lambda_{t,d}}\right), \tag{203}$$

where we used $\beta \gtrsim \sigma\sqrt{d + \log\log T} + 1$ to subsume the first term into the second. Summing the two errors,

$$\left|g_t(\boldsymbol{w}) - \widetilde{g}_t(\boldsymbol{w})\right| = \left|g_{t,1}(\boldsymbol{w}) - \widetilde{g}_{t,1}(\boldsymbol{w})\right| + \left|g_{t,2}(\boldsymbol{w}) - \widetilde{g}_{t,2}(\boldsymbol{w})\right| = O\left(\frac{\beta(\sigma\sqrt{d + \log\log T} + 1)}{\lambda_{t,d}}\right) + O\left(\frac{\beta^2}{\lambda_{t,1}}\right).$$

If, as is typical in sequential designs, $\lambda_{t,1} \gtrsim \lambda_{t,d}$ or even $\lambda_{t,1} \asymp t$, the last term is dominated by the displayed bound (or simplifies to $O(\beta^2/t)$ as in your derivation), leading to the stated rate:

$$\left| g_t(\boldsymbol{w}) - \widetilde{g}_t(\boldsymbol{w}) \right| = O\left(\frac{\beta(\sigma\sqrt{d + \log\log T} + 1)}{\lambda_{t,d}} \right). \tag{204}$$

Suppose that $\boldsymbol{w}_t$ is the maximizer of $g_t(\boldsymbol{w})$ and $\widetilde{\boldsymbol{w}}_t$ is the maximizer of $\widetilde{g}_t(\boldsymbol{w})$. Then it holds that

$$\widetilde{g}_t(\widetilde{\boldsymbol{w}}_t) - \widetilde{g}_t(\boldsymbol{w}_t) = \widetilde{g}_t(\widetilde{\boldsymbol{w}}_t) - g_t(\widetilde{\boldsymbol{w}}_t) + g_t(\widetilde{\boldsymbol{w}}_t) - g_t(\boldsymbol{w}_t) + g_t(\boldsymbol{w}_t) - \widetilde{g}_t(\boldsymbol{w}_t)$$

$$= O\left(\frac{\beta(\sigma\sqrt{d + \log\log T} + 1)}{\lambda_{t,d}} \right), \tag{205}$$

where the last equality holds directly from (204). We will then turn our attention to consider the maximizer of $\widetilde{g}_t(\boldsymbol{w})$. We note that

$$\widetilde{g}_t(\boldsymbol{w}) = 1 + \frac{2\beta w_1}{\sqrt{\lambda_{t,1}}} + \sum_{i=2}^{d} \frac{\beta^2 w_i^2}{\lambda_{t,i}} \leq 1 + \frac{2\beta w_1}{\sqrt{\lambda_{t,1}}} + \frac{\beta^2}{\lambda_{t,d}} \sum_{i=2}^{d} w_i^2$$

$$= 1 + \frac{2\beta w_1}{\sqrt{\lambda_{t,1}}} + \frac{\beta^2}{\lambda_{t,d}}(1 - w_1^2), \tag{206}$$

where the inequality becomes equality if and only if $w_i = 0$ for $2 \leq i \leq d-1$. By directly taking derivative, one note that $\widetilde{g}_t(\boldsymbol{w})$ takes the maximum when

$$\widetilde{w}_{t,1} = \min\left(\frac{\lambda_{t,d}}{\beta\sqrt{\lambda_{t,1}}}, 1 \right) = \frac{\lambda_{t,d}}{\beta\sqrt{\lambda_{t,1}}}, \quad \widetilde{w}_{t,d} = \sqrt{1 - w_1^2}.$$

To characterize $\boldsymbol{w}_t$ with $\widetilde{\boldsymbol{w}}_t$, we first note that

$$\widetilde{g}_t(\widetilde{\boldsymbol{w}}_t) - \widetilde{g}_t(\boldsymbol{w}) = \left(\frac{2\beta\widetilde{w}_{t,1}}{\sqrt{\lambda_{t,1}}} + \frac{\beta^2(1 - \widetilde{w}_{t,1}^2)}{\lambda_{t,d}} \right) - \left(\frac{2\beta w_1}{\sqrt{\lambda_{t,1}}} + \sum_{i=2}^{d} \frac{\beta^2 w_i^2}{\lambda_{t,i}} \right)$$

$$\geq \left(\frac{2\beta\widetilde{w}_{t,1}}{\sqrt{\lambda_{t,1}}} + \frac{\beta^2(1 - \widetilde{w}_{t,1}^2)}{\lambda_{t,d}} \right) - \left(\frac{2\beta w_1}{\sqrt{\lambda_{t,1}}} + \frac{\beta^2(1 - w_1^2)}{\lambda_{t,d}} \right)$$

$$= \frac{2\beta(\widetilde{w}_{t,1} - w_{t,1})}{\sqrt{\lambda_{t,1}}} - \frac{\beta^2(\widetilde{w}_{t,1} - w_{t,1})(\widetilde{w}_{t,1} + w_{t,1})}{\lambda_{t,d}}$$

$$= (\widetilde{w}_{t,1} - w_{t,1}) \cdot \left(\frac{2\beta}{\sqrt{\lambda_{t,1}}} - \frac{\beta^2(\widetilde{w}_{t,1} + w_{t,1})}{\lambda_{t,d}} \right). \tag{207}$$

Here, one can note that when $\widetilde{w}_{t,1} = \frac{\lambda_{t,d}}{\beta\sqrt{\lambda_{t,1}}}$, the difference can be rewritten as

$$\widetilde{g}_t(\widetilde{\boldsymbol{w}}_t) - \widetilde{g}_t(\boldsymbol{w}) \geq \frac{\beta^2(\widetilde{w}_{t,1} - w_{t,1})^2}{\lambda_{t,d}}, \tag{208}$$

and when $\widetilde{w}_{t,1} = 1$, the difference can be rewritten as

$$\widetilde{g}_t(\widetilde{\boldsymbol{w}}_t) - \widetilde{g}_t(\boldsymbol{w}) \geq (\widetilde{w}_{t,1} - w_{t,1}) \cdot \left(\frac{2\beta}{\sqrt{\lambda_{t,1}}} - \frac{\beta^2}{\lambda_{t,d}} - \frac{\beta^2 w_{t,1}}{\lambda_{t,d}} \right)$$

$$\geq \frac{\beta^2(\widetilde{w}_{t,1} - w_{t,1})^2}{\lambda_{t,d}}, \tag{209}$$

where the last inequality holds because $\lambda_{t,d}/(\beta\sqrt{\lambda_{t,1}}) \geq 1$. As we require that

$$\widetilde{g}_t(\widetilde{\boldsymbol{w}}_t) - \widetilde{g}_t(\boldsymbol{w}_t) = O\left(\frac{\beta(\sigma\sqrt{d + \log\log T} + 1)}{\lambda_{t,d}} \right), \tag{210}$$

the difference between $\widetilde{w}_{t,1}$ and $w_{t,1}$ can be bounded as

$$|\widetilde{w}_{t,1} - w_{t,1}| = O\left(\frac{(\sigma\sqrt{d + \log\log T} + 1)^{1/2}}{\sqrt{\beta}}\right). \tag{211}$$

From this, the precise expression of $w_{t,1}$ is given as

$$
\begin{aligned}
w_{t,1} &= \widetilde{w}_{t,1} + O\left(\frac{(\sigma\sqrt{d + \log\log T} + 1)^{1/2}}{\sqrt{\beta}}\right) \\
&= \min\left(\frac{\lambda_{t,d}}{\beta\sqrt{\lambda_{t,1}}}, 1\right) + O\left(\frac{(\sigma\sqrt{d + \log\log T} + 1)^{1/2}}{\sqrt{\beta}}\right) \\
&= \min\left[\frac{\lambda_{t,d}}{\beta\sqrt{t}} \cdot \left(\sqrt{\frac{t}{\lambda_{t,1}}} - 1 + 1\right), 1\right] + O\left(\frac{(\sigma\sqrt{d + \log\log T} + 1)^{1/2}}{\sqrt{\beta}}\right) \\
&= \min\left[\frac{\lambda_{t,d}}{\beta\sqrt{t}}\left(1 + O\left(\frac{\beta\sqrt{d}}{\sqrt{t}}\right)\right), 1\right] + O\left(\frac{(\sigma\sqrt{d + \log\log T} + 1)^{1/2}}{\sqrt{\beta}}\right) \\
&= \min\left(\frac{\lambda_{t,d}}{\beta\sqrt{t}}, 1\right) + O\left(\frac{(\sigma\sqrt{d + \log\log T} + 1)^{1/2}}{\sqrt{\beta}}\right),
\end{aligned} \tag{212}
$$

where the penultimate line holds as

$$\sqrt{\frac{t}{\lambda_{t,1}}} - 1 = \sqrt{\frac{t}{t - (d-1)\bar{\lambda}_t}} = \sqrt{\frac{t}{t - O(\beta\sqrt{d}t)}} = \sqrt{\frac{1}{1 - O(\beta\sqrt{d}/\sqrt{t})}} = 1 + O\left(\frac{\beta\sqrt{d}}{\sqrt{t}}\right). \tag{213}$$

and the last inequality holds when $t \gtrsim \beta^3/(\sigma\sqrt{d})$,

$$\frac{\lambda_{t,d}}{\beta\sqrt{t}} \cdot \frac{\beta\sqrt{d}}{\sqrt{t}} = \frac{\lambda_{t,d}\sqrt{d}}{t} = O\left(\frac{\beta}{\sqrt{t}}\right) = O\left(\frac{(\sigma\sqrt{d + \log\log T} + 1)^{1/2}}{\sqrt{\beta}}\right) \tag{214}$$

We also analyze behavior of $w_{t,i}$ for $i \geq 2$. Note that

$$
\begin{aligned}
\widetilde{g}_t(\widetilde{\boldsymbol{w}}_t) - \widetilde{g}_t(\boldsymbol{w}_t) &= \left(\frac{2\beta\widetilde{w}_{t,1}}{\sqrt{\lambda_{t,1}}} + \frac{\beta^2(1 - \widetilde{w}_{t,1}^2)}{\lambda_{t,d}}\right) - \left(\frac{2\beta w_{t,1}}{\sqrt{\lambda_{t,1}}} + \sum_{i=2}^{d}\frac{\beta^2 w_{t,i}^2}{\lambda_{t,i}}\right) \\
&= \left(\frac{2\beta\widetilde{w}_{t,1}}{\sqrt{\lambda_{t,1}}} + \frac{\beta^2(1 - \widetilde{w}_{t,1}^2)}{\lambda_{t,d}}\right) - \left(\frac{2\beta w_{t,1}}{\sqrt{\lambda_{t,1}}} + \frac{\beta^2(1 - w_{t,1}^2)}{\lambda_{t,d}}\right) \\
&\quad + \left(\sum_{i=2}^{d}\frac{\beta^2 w_{t,i}^2}{\lambda_{t,d}} - \sum_{i=2}^{d}\frac{\beta^2 w_{t,i}^2}{\lambda_{t,i}}\right) \\
&\geq \beta^2 \sum_{i=2}^{d}\left(\frac{w_{t,i}^2}{\lambda_{t,d}} - \frac{w_{t,i}^2}{\lambda_{t,i}}\right).
\end{aligned} \tag{215}
$$

As a result, one can show that

$$\sum_{i=2}^{d}\left(\frac{w_{t,i}^2}{\lambda_{t,d}} - \frac{w_{t,i}^2}{\lambda_{t,i}}\right) = O\left(\frac{\sigma\sqrt{d + \log\log T} + 1}{\beta\lambda_{t,d}}\right). \tag{216}$$

From (178), it satisfies that for any $\|\boldsymbol{w}\|_2 = 1$, it holds uniformly that

$$g_t(\boldsymbol{w}) = 1 + O\left(\frac{\beta^2}{\lambda_{t,d}}\right). \tag{217}$$

56

Therefore, the action vector $\boldsymbol{a}_t$ can be characterized as

$$\boldsymbol{a}_t = [g_t(\boldsymbol{w}_t)]^{-1} \cdot (\widehat{\boldsymbol{\theta}}_t + \beta \boldsymbol{\Lambda}_t^{-1/2} \boldsymbol{w}) = \left(1 - O\left(\frac{\beta^2}{\lambda_{t,d}}\right)\right) \cdot \sum_{i=1}^{d} \left(\nu_{t,i} + \frac{\beta w_{t,i}}{\sqrt{\lambda_{t,i}}}\right) \boldsymbol{v}_{t,i}. \tag{218}$$

Consequently, one has

$$\sum_{i=2}^{d} \kappa_{t,i}^2 = \left(1 - O\left(\frac{\beta^2}{\lambda_{t,d}}\right)\right) \cdot \left(\sum_{i=2}^{d} \frac{\beta^2 w_{t,i}^2}{\lambda_{t,i}} + \sum_{i=2}^{d} \frac{2\beta \nu_{t,i} w_{t,i}}{\sqrt{\lambda_{t,i}}} + \sum_{i=2}^{d} \nu_{t,i}^2\right)$$

$$= \left(1 - O\left(\frac{\beta^2}{\lambda_{t,d}}\right)\right) \cdot \left(\beta^2 \left(\sum_{i=2}^{d} \frac{w_{t,i}^2}{\lambda_{t,d}} - O\left(\frac{\sigma\sqrt{d + \log\log T} + 1}{\beta \lambda_{t,d}}\right)\right) + 2\beta \sqrt{\sum_{i=2}^{d} \frac{w_{t,i}^2}{\lambda_{t,i}}} \sqrt{\sum_{i=2}^{d} \nu_{t,i}^2} + \sum_{i=2}^{d} \nu_{t,i}^2\right)$$

$$= \left(1 - O\left(\frac{\beta^2}{\lambda_{t,d}}\right)\right) \cdot \left(\beta^2 \left(\frac{1 - w_{t,1}^2}{\lambda_{t,d}} - O\left(\frac{\sigma\sqrt{d + \log\log T} + 1}{\beta \lambda_{t,d}}\right)\right)\right.$$

$$\left. + 2\beta \cdot \sqrt{\frac{1 - w_{t,1}^2}{\lambda_{t,d}}} \cdot \frac{\sigma\sqrt{d + \log\log T} + 1}{\sqrt{\lambda_{t,d}}} + \frac{(\sigma\sqrt{d + \log\log T} + 1)^2}{\lambda_{t,d}}\right)$$

$$= \left(1 - O\left(\frac{\beta^2}{\lambda_{t,d}}\right)\right) \cdot \beta^2 \left(\frac{1 - w_{t,1}^2}{\lambda_{t,d}} - O\left(\frac{\sigma\sqrt{d + \log\log T} + 1}{\beta \lambda_{t,d}}\right)\right)$$

$$= \frac{\beta^2}{\lambda_{t,d}} \left(1 - w_{t,1}^2 + O\left(\frac{\sigma\sqrt{d + \log\log T} + 1}{\beta}\right)\right). \tag{219}$$

where the last equality holds whenever $t \gtrsim \beta^4 d / \sigma^2$. Plugging in the expression of $w_{t,1}$ in (212), we obtain that

$$\sum_{i=2}^{d} \kappa_{t,i}^2 = \frac{\beta^2}{\lambda_{t,d}} \left(1 - \left(\widetilde{w}_{t,1} + O\left(\frac{(\sigma\sqrt{d + \log\log T} + 1)^{1/2}}{\sqrt{\beta}}\right)\right)^2 + O\left(\frac{\sigma\sqrt{d + \log\log T} + 1}{\beta}\right)\right)$$

$$= \frac{\beta^2}{\lambda_{t,d}} \left(1 - \widetilde{w}_{t,1}^2 + O\left(\frac{(\sigma\sqrt{d + \log\log T} + 1)^{1/2}}{\sqrt{\beta}}\right)\right). \tag{220}$$

Step 2: preliminary growth speed control of $\lambda_{t,d}$. The above decomposition is the key to characterize the growth and concentration of non-leading eigenvalues. However, we note that the growth of non-leading eigenvalues could not be precisely characterized when $\lambda_{t,d}$ is large. Therefore, before we establish any more fine-grained result, we first need to show that for some $c < 1$, there exists t_2' such that for all $t \geq t_2'$, the minimum eigenvalue is controlled as $\lambda_{t,d} \leq c\beta\sqrt{t}$.

We begin from Lemma 5, one can show that

$$\lambda_{t+1,1} \geq \lambda_{t,1} + \kappa_{t,1}^2,$$

which implies that

$$\overline{\lambda}_T \leq \overline{\lambda}_t + \frac{\sum_{i=2}^{d} \kappa_{t,i}^2}{d - 1}. \tag{221}$$

Therefore, whenever $\lambda_{t,d} > c\beta\sqrt{t}$, we can lower bound $w_{t,1}$ as

$$w_{t,1} \geq c + O\left(\frac{(\sigma\sqrt{d + \log\log T} + 1)^{1/2}}{\sqrt{\beta}}\right), \tag{222}$$

which allows us to upper bound the growth of $\overline{\lambda}_t$ as

$$\overline{\lambda}_T \leq \overline{\lambda}_t + \frac{1}{d - 1} \cdot \frac{\beta^2}{\lambda_{t,d}} \left(1 - w_{t,1}^2 + O\left(\frac{\sigma\sqrt{d + \log\log T} + 1}{\beta}\right)\right)$$

$$= \overline{\lambda}_t + \frac{1}{d-1} \cdot \frac{\beta^2}{\lambda_{t,d}} \left(1 - c^2 + O\left(\frac{(\sigma\sqrt{d + \log\log T} + 1)^{1/2}}{\sqrt{\beta}} \right) \right)$$

$$= \overline{\lambda}_t + \frac{1}{d-1} \cdot \frac{\beta}{c\sqrt{t}} \left(1 - c^2 + O\left(\frac{(\sigma\sqrt{d + \log\log T} + 1)^{1/2}}{\sqrt{\beta}} \right) \right)$$

$$= \overline{\lambda}_t + \frac{\overline{c}\beta}{\sqrt{t}}, \tag{223}$$

where $\overline{c}$ is defined as

$$\overline{c} = \frac{1}{c(d-1)} \left(1 - c^2 + O\left(\frac{(\sigma\sqrt{d + \log\log T} + 1)^{1/2}}{\sqrt{\beta}} \right) \right).$$

One can show that when T is large enough, $d \geq 2$ and c is chosen close enough to 1, it holds that

$$\overline{c} \leq \frac{1 - \frac{3}{4}c^2}{c} \leq \frac{c}{3}, \tag{224}$$

implying that for any $t \geq t_2$, and that $\lambda_{t,d} \geq c\beta\sqrt{t}$ holds for any time indices between t_2 and t, it holds that

$$\overline{\lambda}_t - \overline{\lambda}_{t_2} \leq \sum_{t'=t_2}^{t} \frac{\overline{c}\beta}{\sqrt{t'}} \leq \frac{c\beta}{3} \sum_{t'=t_2}^{t} \frac{1}{\sqrt{t'}} \leq \frac{2c\beta}{3}(\sqrt{t} - \sqrt{t_2}). \tag{225}$$

As $\overline{\lambda}_{t_2} = O(\beta\sqrt{t_2})$, with this strategy, one can show that there exists $t_2' = O(t_2)$, such that $\overline{\lambda}_{t_2'} \leq c\beta\sqrt{t_2'}$. We will then show that for any $t \geq t_2'$, it holds that $\overline{\lambda}_t \leq c\beta\sqrt{t}$. Suppose that this does not always hold, this implies that there exists t such that $\overline{\lambda}_t \leq c\beta\sqrt{t}$ but $\overline{\lambda}_T > c\beta\sqrt{t+1}$. Since $\overline{\lambda}_T - \overline{\lambda}_t \leq 1$, we can assume that

$$\overline{\lambda}_t \geq \frac{99}{100}c\beta\sqrt{t}.$$

Then it holds that

$$\overline{\lambda}_T - \overline{\lambda}_t \leq \frac{3c\beta}{8\sqrt{t}} < c\beta(\sqrt{t+1} - \sqrt{t}), \tag{226}$$

which leads to a contradiction! Therefore, we conclude that for any $c < 1$, there exists $t_2' = O(t_2)$ such that for all $t \geq t_2$, $\lambda_{t,d} \leq \overline{\lambda}_t \leq c\beta\sqrt{t}$.

Step 3: limiting the projection of of a_t on eigenspace of large eigenvalues. After the growth smallest eigenvalue is being controlled, we can then precisely characterize the growth of non-leading eigenvalues and then upper bound the projection of a_t on the large components. To proceed, we adapt a similar strategy that was used in Phase #1 and Phase #2. We construct a set of large eigenvalues as

$$\mathcal{L}_t = \left\{ i : \ i \geq 2, \ \lambda_{t,i} \geq \left(1 + C_1 \cdot \frac{d(\sigma\sqrt{d + \log\log T} + 1)}{\beta} \right) \lambda_{t,d} \right\}.$$

Then one can upper bound the summation of $w_{t,i}^2$ within the set of large eigenvalues in the following way

$$\sum_{i=2}^{d} \left(\frac{w_{t,i}^2}{\lambda_{t,d}} - \frac{w_{t,i}^2}{\lambda_{t,i}} \right) \geq \sum_{i\in\mathcal{L}_t} \left(\frac{w_{t,i}^2}{\lambda_{t,d}} - \frac{w_{t,i}^2}{\lambda_{t,i}} \right) = \frac{1}{\lambda_{t,d}} \sum_{i\in\mathcal{L}_t} w_{t,i}^2 \left(1 - \frac{\lambda_{t,d}}{\lambda_{t,i}} \right) \geq \frac{C_1 d(\sigma\sqrt{d + \log\log T} + 1)}{2\beta\lambda_{t,d}} \cdot \sum_{i\in\mathcal{L}_t} w_{t,i}^2, \tag{227}$$

which holds as $\beta \geq C(\sigma\sqrt{d + \log\log T} + 1)$ for some constant C. Here, as we set C_1 large enough, we have

$$\sum_{i\in\mathcal{L}_t} w_{t,i}^2 \leq \frac{1}{8d}(1 - c^2). \tag{228}$$

58

We can then upper bound the summation of $\kappa_{t,i}^2$ within the subset as

$$\sum_{i \in \mathcal{L}_t} \kappa_{t,i}^2 \le \sum_{i \in \mathcal{L}_t} \left(\nu_{t,i} + \frac{\beta w_{t,i}}{\sqrt{\lambda_{t,i}}} \right)^2 \le \left(\sqrt{\sum_{i \in \mathcal{L}_t} \nu_{t,i}^2} + \sqrt{\sum_{i \in \mathcal{L}_t} \frac{\beta^2 w_{t,i}^2}{\lambda_{t,i}}} \right)^2$$

$$= \left[O\left(\frac{\sigma\sqrt{d + \log\log T} + 1}{\sqrt{\lambda_{t,d}}} \right) + \frac{\beta}{\sqrt{\lambda_{t,d}}} \cdot \sqrt{\sum_{i \in \mathcal{L}_t} w_{t,i}^2} \right]^2$$

$$= \frac{\beta^2}{\lambda_{t,d}} \left(\sqrt{\sum_{i \in \mathcal{L}_t} w_{t,i}^2} + O\left(\frac{\sigma\sqrt{d + \log\log T} + 1}{\beta} \right) \right)^2$$

$$= \frac{\beta^2}{\lambda_{t,d}} \left(\sum_{i \in \mathcal{L}_t} w_{t,i}^2 + O\left(\frac{\sigma\sqrt{d + \log\log T} + 1}{\beta} \right) \right), \tag{229}$$

and it holds that when $\lambda_{t,d} \le c\beta\sqrt{t}$,

$$\sum_{i \in \mathcal{L}_t} \kappa_{t,i}^2 \le \frac{\sum_{i \in \mathcal{L}_t} w_{t,i}^2 + O\left(\frac{\sigma\sqrt{d + \log\log T} + 1}{\beta} \right)}{1 - w_{t,1}^2 + O\left(\frac{\sigma\sqrt{d + \log\log T} + 1}{\beta} \right)} \cdot \sum_{i=2}^d \kappa_{t,i}^2$$

$$= \frac{\sum_{i \in \mathcal{L}_t} w_{t,i}^2 + O\left(\frac{\sigma\sqrt{d + \log\log T} + 1}{\beta} \right)}{1 - \frac{\lambda_{t,d}^2}{\beta^2 t} + O\left(\frac{(\sigma\sqrt{d + \log\log T} + 1)^{1/2}}{\sqrt{\beta}} \right)} \cdot \sum_{i=2}^d \kappa_{t,i}^2$$

$$\le \frac{\frac{1}{6d}(1 - c^2)}{\frac{2}{3}(1 - c^2)} \cdot \sum_{i=2}^d \kappa_{t,i}^2$$

$$\le \frac{1}{4d} \sum_{i=2}^d \kappa_{t,i}^2. \tag{230}$$

Step 4: limiting the growth of large non-leading eigenvalues. Equipped an upper bound of the projection of $\boldsymbol{a}_t$ on the eigenspace of large non-leading eigenvalues, we will then bound the growth speed of the large non-leading eigenvalues, therefore establishing the concentration of all non-leading eigenvalues by the end of Phase #4. To show the final concentration, we only need to show the concentration of $\lambda_{t,2}$ and $\lambda_{t,d}$, i.e. $\lambda_{t,2}/\lambda_{t,d} = 1 + o(1)$.

We aim to show that the second-largest eigenvalue $\lambda_{t,2}$ cannot grow significantly faster than the rest of the non-leading spectrum—more specifically, it cannot stay much larger than the average $\lambda_{t,d}$ for an extended period. To formalize this, we introduce a higher threshold for non-leading eigenvalues,

$$\widetilde{\mathcal{L}}_t = \left\{ i : \ i \ge 2, \ \lambda_{t,i} \ge \left(1 + 2C_1 \cdot \frac{d(\sigma\sqrt{d + \log\log T} + 1)}{\beta} \right) \lambda_{t,d} \right\}.$$

We will demonstrate that such a deviation is not sustainable over time by analyzing the eigenvalue update induced by the rank-one perturbation at each step. To proceed, recall the secular function for the rank-one update:

$$f(\lambda) = 1 + \sum_{i=1}^d \frac{\kappa_{t,i}^2}{\lambda_{t,i} - \lambda}.$$

This function determines the characteristic equation for the updated eigenvalues at time $t + 1$. When $\lambda_{t,2} \in \widetilde{\mathcal{L}}_t$, consider a candidate point $\widetilde{\lambda}_{t+1,2}$ satisfying

$$\widetilde{\lambda}_{t+1,2} = \lambda_{t,2} + \frac{1}{2d} \sum_{i=2}^d \kappa_{t,i}^2,$$

59

and claim that $\lambda_{t+1,2} \leq \widetilde{\lambda}_{t+1,2}$. To show this result, we only need to show that $f(\widetilde{\lambda}_{t+1,2}) > 0$. We analyze $f(\lambda)$ to determine whether this hold true. Here, we denote:

$$\delta = C_1 \cdot \frac{d(\sigma\sqrt{d + \log\log T} + 1)}{\beta}.$$

We partition the summation in $f(\lambda)$ based on whether the denominator $\lambda_{t,i} - \lambda$ is relatively small or large. Specifically, split the indices into three groups:

$$f(\widetilde{\lambda}_{t+1,2}) = 1 + \frac{\kappa_{t,1}^2}{\lambda_{t,1} - \widetilde{\lambda}_{t+1,2}} + \sum_{i:\ i \geq 2, \lambda_{t,i} \geq \lambda_{t,2} - \delta\lambda_{t,d}} \frac{\kappa_{t,i}^2}{\lambda_{t,i} - \widetilde{\lambda}_{t+1,2}} + \sum_{i:\lambda_{t,i} < \lambda_{t,2} - \delta\lambda_{t,d}} \frac{\kappa_{t,i}^2}{\lambda_{t,i} - \widetilde{\lambda}_{t+1,2}}. \tag{231}$$

For the leading eigenvalue $\lambda_{t,1}$, as we have shown in Proposition 1 that for any $t \geq t_1$, it holds that $\lambda_{t,1} \geq 8t/15$. Therefore, we conclude that when $\lambda_{t,1} - \widetilde{\lambda}_{t+1,2} \geq t/20$. Therefore, one has

$$\frac{\kappa_{t,1}^2}{\lambda_{t,1} - \widetilde{\lambda}_{t+1,2}} = O(t^{-1}). \tag{232}$$

For the group of "large" non-leading eigenvalues (where $\lambda_{t,i}$ is close to $\lambda_{t,2}$), note that

$$\widetilde{\lambda}_{t+1,2} - \lambda_{t,i} \geq \widetilde{\lambda}_{t+1,2} - \lambda_{t,2} \geq \frac{1}{2d}\sum_{j=2}^{d}\kappa_{t,j}^2,$$

hence one has the following lower bound

$$\frac{\kappa_{t,i}^2}{\lambda_{t,i} - \lambda} \geq -\kappa_{t,i}^2 \cdot \frac{2d}{\sum_{j=2}^{d}\kappa_{t,j}^2}. \tag{233}$$

For the group of "small eigenvalues" (where $\lambda_{t,i}$ is much smaller), we use the assumption that $\lambda_{t,i} - \lambda < -\delta\lambda_{t,d}$, leading to the bound

$$\frac{\kappa_{t,i}^2}{\lambda_{t,i} - \lambda} \geq -\frac{\kappa_{t,i}^2}{\delta\lambda_{t,d}}. \tag{234}$$

Combining and simplifying terms yields the following lower bound on $f(\lambda)$:

$$\begin{aligned}
f(\lambda) &\geq 1 + O(t^{-1}) - \sum_{i:\lambda_{t,i} \geq \lambda_{t,2} - \delta\lambda_{t,d}} \kappa_{t,i}^2 \cdot \frac{2d}{\sum_{j=2}^{d}\kappa_{t,j}^2} - \sum_{i:\lambda_{t,i} < \lambda_{t,2} - \delta\lambda_{t,d}} \kappa_{t,i}^2 \cdot \frac{1}{\delta\lambda_{t,d}} \\
&\geq 1 + O(t^{-1}) - \sum_{i \in \mathcal{L}_t} \kappa_{t,i}^2 \cdot \frac{2d}{\sum_{j=2}^{d}\kappa_{t,j}^2} - \left(\sum_{i=2}^{d}\kappa_{t,i}^2 - \sum_{i \in \mathcal{L}_t}\kappa_{t,i}^2\right) \cdot \frac{1}{\delta\lambda_{t,d}} \\
&\geq 1 + O(t^{-1}) - \frac{\sum_{i=2}^{d}\kappa_{t,i}^2}{4d} \cdot \frac{2d}{\sum_{i=2}^{d}\kappa_{t,i}^2} - \left(1 - \frac{1}{4d}\right) \cdot \frac{\sum_{i=2}^{d}\kappa_{t,i}^2}{\delta\lambda_{t,d}} \\
&\geq \frac{1}{2} + O(t^{-1}) - \frac{\sum_{i=2}^{d}\kappa_{t,i}^2}{\delta\lambda_{t,d}} > 0,
\end{aligned} \tag{235}$$

where the last inequality holds for $t \geq t_2$, since

$$\frac{\sum_{i=2}^{d}\kappa_{t,i}^2}{\delta\lambda_{t,d}} \lesssim \frac{\beta^2}{\lambda_{t,d}} \cdot \frac{1}{\delta\lambda_{t,d}} \lesssim \frac{\beta^3}{\lambda_{t,d}^2 d} = \frac{\beta}{t}.$$

The positivity of $f(\widetilde{\lambda}_{t+1,2})$ implies that the updated new eigenvalue $\lambda_{t+1,2}$ can be upper bounded by $\widetilde{\lambda}_{t+1,2}$, hence we have

$$\lambda_{t+1,2} \leq \lambda_{t,2} + \frac{1}{2d}\sum_{i=2}^{d}\kappa_{t,i}^2, \tag{236}$$

whenever $\lambda_{t,2} \geq (1 + 2\delta)\lambda_{t,d}$. In other words, the second eigenvalue cannot grow faster than this speed if it deviates much from the minimum eigenvalue. On the other hand, from Lemma 6, the sum of non-leading eigenvalues evolves as

$$\sum_{i=2}^{d} \lambda_{t+1,i} = \sum_{i=2}^{d} \lambda_{t,i} + \sum_{i=2}^{d} \kappa_{t,i}^2 + O(t^{-1}),$$

which implies the average satisfies

$$\overline{\lambda}_T = \overline{\lambda}_t + \frac{1}{d-1} \sum_{i=2}^{d} \kappa_{t,i}^2 + O(t^{-1}) \geq \overline{\lambda}_t + \frac{1}{d} \sum_{i=2}^{d} \kappa_{t,i}^2. \tag{237}$$

Combining the bounds on the growth of $\lambda_{t,2}$ and $\overline{\lambda}_t$, we obtain that whenever $\lambda_{t,2} \geq (1 + 2\delta)\lambda_{t,d}$,

$$\lambda_{t+1,2} - \lambda_{t,2} \leq \frac{1}{2d} \sum_{i=2}^{d} \kappa_{t,i}^2 \leq \frac{1}{2}(\overline{\lambda}_T - \overline{\lambda}_t). \tag{238}$$

This inequality is crucial: it shows that whenever $\lambda_{t,2}$ becomes disproportionately large relative to $\lambda_{t,d}$, its future growth is outpaced by the average $\overline{\lambda}_t$. Therefore, any such deviation must shrink over time.

To show the desired result, we define the set $\mathcal{S} = \{t \geq t_2 : \lambda_{t,2} \geq (1 + 2\delta)\lambda_{t,d}\}$, which collects the time indices where the second-largest eigenvalue is significantly larger than the smallest non-leading eigenvalue. Intuitively, this set captures the time intervals when the spectrum is relatively "spread out." Due to the nature of eigenvalue evolution, the set $\mathcal{S}$ may consist of multiple disjoint time intervals or "segments"—each comprising consecutive time steps during which the elevated second eigenvalue persists.

To understand the long-term behavior of $\lambda_{t,2}$, we analyze the possible structure of $\mathcal{S}$ by examining different types of segments. In particular, we consider two cases: the initial segment (beginning at t_2) and general intermediate segments where the elevated condition temporarily reappears.

Case 1: initial segment of $\mathcal{S}$. Suppose the initial point $t_2 \in \mathcal{S}$, and that the elevated condition $\lambda_{t,2} \geq (1 + 2\delta)\lambda_{t,d}$ holds throughout the interval $[t_2, t']$. Applying the growth inequality established earlier:

$$\lambda_{t+1,2} - \lambda_{t,2} \leq \frac{1}{2d} \sum_{i=2}^{d} \kappa_{t,i}^2 \leq \frac{1}{2}(\overline{\lambda}_T - \overline{\lambda}_t),$$

and telescoping over the interval $[t_2', t']$, we obtain

$$\lambda_{t',2} - \lambda_{t_2,2} \leq \frac{1}{2}(\overline{\lambda}_{t'} - \overline{\lambda}_{t_2}).$$

Rearranging gives

$$\begin{aligned}
\lambda_{t',2} - \overline{\lambda}_{t'} &\leq \lambda_{t',2} - \lambda_{t_2,2} + \lambda_{t_2,2} - \overline{\lambda}_{t'} \\
&\leq \frac{1}{2}(\overline{\lambda}_{t'} - \overline{\lambda}_{t_2}) + \lambda_{t_2,2} - \overline{\lambda}_{t'} \\
&= (\lambda_{t_2,2} - \overline{\lambda}_{t_2}) - \frac{1}{2}(\overline{\lambda}_{t'} - \overline{\lambda}_{t_2}).
\end{aligned}$$

This shows that the deviation $\lambda_{t',2} - \overline{\lambda}_{t'}$ decreases unless the growth of $\lambda_{t',2}$ is concentrated close to $\lambda_{t,d}$

If the right-hand side is positive, we can bound $\overline{\lambda}_{t'}$ and $\lambda_{t',2}$ as follows:

$$\overline{\lambda}_{t'} \leq \overline{\lambda}_{t_2} + 2(\lambda_{t_2,2} - \overline{\lambda}_{t_2}),$$

$$\lambda_{t',2} \leq (\lambda_{t_2,2} - \overline{\lambda}_{t_2}) + \frac{1}{2}(\overline{\lambda}_{t'} + \overline{\lambda}_{t_2}) \leq \lambda_{t_2,2} + (\lambda_{t_2,2} - \overline{\lambda}_{t_2}).$$

Since $\lambda_{t_2,2} \leq d\overline{\lambda}_{t_2}$, it follows that $\lambda_{t',2} = O(\beta\sqrt{dt_2})$. Meanwhile, the average $\overline{\lambda}_{t'}$ continues to grow roughly as $\Omega(\beta\sqrt{t'}/\sqrt{d})$, and hence for sufficiently large C, we get a contradiction to the assumption $\lambda_{t',2} \geq (1 + 2\delta)\lambda_{t',d} \approx \Omega(\overline{\lambda}_{t'})$. Thus, this shows that the initial segment of $\mathcal{S}$ cannot persist for too long and must terminate by some time $t_4 = O(t_2'd^2) = O(\beta^8/\sigma^6)$, where $t_4 \notin \mathcal{S}$.

Case 2: intermediate segments of $\mathcal{S}$. Now suppose that $\mathcal{S}$ reappears after being interrupted—specifically, suppose there exists a time $t'' \notin \mathcal{S}$ such that $t'' - 1 \in \mathcal{S}$, i.e., the second eigenvalue has just dropped below the elevated threshold. Then, by continuity of the update dynamics and the rank-one perturbation nature of the process, the spectral gap at time t'' is bounded:

$$\lambda_{t'',2} - \lambda_{t'',d} \leq \lambda_{t''-1,2} - \lambda_{t''-1,d} + 1 \leq 2\delta\lambda_{t'',d} + 1 \leq 3\delta\lambda_{t'',d}.$$

Now suppose the elevated condition re-emerges and persists from t'' to some later time t' (i.e., $t \in \mathcal{S}$ for all $t'' \leq t \leq t'$). Telescoping again over this interval yields

$$\lambda_{t',2} - \lambda_{t'',2} \leq \frac{1}{2}(\overline{\lambda}_{t'} - \overline{\lambda}_{t''}),$$

which implies

$$\lambda_{t',2} - \overline{\lambda}_{t'} \leq (\lambda_{t_2,2} - \overline{\lambda}_{t''}) - \frac{1}{2}(\overline{\lambda}_{t'} - \overline{\lambda}_{t''}) \leq (\lambda_{t'',2} - \overline{\lambda}_{t''}) - (\lambda_{t',2} - \lambda_{t'',2}).$$

This shows that the deviation from the average remains controlled. To bound the total gap $\lambda_{t',2} - \lambda_{t',d}$, we use the fact that both the shift from $\lambda_{t'',2}$ and the previous deviation from $\lambda_{t'',d}$ are bounded:

$$\begin{aligned}
\lambda_{t',2} - \lambda_{t',d} &\leq \min\left(\lambda_{t',2} - \lambda_{t'',2} + \lambda_{t'',2} - \lambda_{t'',d}, \ d(\lambda_{t',2} - \overline{\lambda}_{t'})\right) \\
&= \min\left(\lambda_{t',2} - \lambda_{t'',2} + 3\delta\lambda_{t'',d}, \ 3\delta d\lambda_{t'',d} - d(\lambda_{t',2} - \lambda_{t'',2})\right) \\
&\leq 6\delta\lambda_{t'',d} \leq 6\delta\lambda_{t',d}.
\end{aligned} \tag{239}$$

Hence, even if the elevated condition reappears in later intervals, the gap between the second and smallest eigenvalues remains proportionally bounded. In particular, for all $t \geq t_4$, we have

$$\frac{\lambda_{t,2} - \lambda_{t,d}}{\lambda_{t,d}} \leq 6\delta,$$

which ensures that $\lambda_{t,2}$ cannot significantly exceed the rest of the spectrum in the long run.

Finally, since for any $i \geq 2$, $\lambda_{t,i} \leq \lambda_{t,2}$ and $\overline{\lambda}_t \geq \lambda_{t,d}$, we obtain

$$\frac{\lambda_{t,i} - \overline{\lambda}_t}{\overline{\lambda}_t} \leq \frac{\lambda_{t,2} - \lambda_{t,d}}{\lambda_{t,d}} = O\left(\frac{d(\sigma\sqrt{d + \log\log T} + 1)}{\beta}\right), \tag{240}$$

as desired.

Step 5: a precise characterization on non-leading eigenvalues. Finally, we give a precise characterization on $\lambda_{t,i}$ when $t \geq t_3$. To show this result, we will first precisely characterize the growth of non-leading eigenvalues. From Lemma 6, one can characterize the growth of non-leading eigenvalues as follows,

$$\sum_{i=2}^{d} \lambda_{t+1,i} = \sum_{i=2}^{d} \lambda_{t,i} + \sum_{i=2}^{d} \kappa_{t,i}^2 + O(t^{-1}), \tag{241}$$

which is equivalent to

$$\begin{aligned}
\overline{\lambda}_T &= \overline{\lambda}_t + \frac{1}{d-1}\sum_{i=2}^{d} \kappa_{t,i}^2 + O(t^{-1}) \\
&= \overline{\lambda}_t + \frac{1}{d-1} \cdot \frac{\beta^2}{\lambda_{t,d}}\left(1 - \frac{\lambda_{t,d}^2}{\beta^2 t} + O\left(\frac{(\sigma\sqrt{d+\log\log T}+1)^{1/2}}{\sqrt{\beta}}\right)\right) + O(t^{-1}) \\
&= \overline{\lambda}_t + \frac{1}{d-1} \cdot \frac{\beta^2}{\overline{\lambda}_t}\left[1 - \frac{\overline{\lambda}_t^2}{\beta^2 t} + O\left(\frac{(\sigma\sqrt{d+\log\log T}+1)^{1/2}}{\sqrt{\beta}}\right)\right] \cdot \left[1 + O\left(\frac{d(\sigma\sqrt{d+\log\log T}+1)}{\beta}\right)\right]
\end{aligned}$$

$$= \overline{\lambda}_t + \frac{1}{d-1} \cdot \frac{\beta^2}{\overline{\lambda}_t} \left[1 - \frac{\overline{\lambda}_t^2}{\beta^2 t} + O\left(\frac{(\sigma\sqrt{d + \log\log T} + 1)^{1/2}}{\sqrt{\beta}} \right) \right], \tag{242}$$

where the last equality holds whenever $\beta \gtrsim d^2(\sigma\sqrt{d + \log\log T} + 1)$. Taking squares on both sides of (242) yields

$$\overline{\lambda}_T^2 = \overline{\lambda}_t^2 + \frac{2\beta^2}{d-1}\left(1 - \frac{\overline{\lambda}_t^2}{\beta^2 t}\right) + O\left(\frac{\beta^2(\sigma\sqrt{d + \log\log T} + 1)^{1/2}}{\sqrt{\beta}(d-1)} \right).$$

Setting $b_t = \frac{\overline{\lambda}_t^2}{\beta^2 t}$, we can write

$$(t+1)b_T = tb_t + \frac{2}{d-1}(1 - b_t) + O\left(\frac{(\sigma\sqrt{d + \log\log T} + 1)^{1/2}}{\sqrt{\beta}d} \right). \tag{243}$$

Define $b_\star = \frac{2}{d+1}$, then we note that $b_\star$ satisfies

$$(t+1)b_\star = tb_\star + \frac{2}{d-1}(1 - b_\star).$$

As a result, set $\overline{\Delta}_t = b_t - b_\star$, (243) can be expressed as

$$(t+1)\overline{\Delta}_T = \left(t - \frac{2}{d-1} \right)\overline{\Delta}_t + O\left(\frac{(\sigma\sqrt{d + \log\log T} + 1)^{1/2}}{\sqrt{\beta}d} \right),$$

which yields the following induction for $|\overline{\Delta}_t|$,

$$\overline{\Delta}_T = \frac{t - \frac{2}{d-1}}{t+1}\overline{\Delta}_t + O\left(\frac{(\sigma\sqrt{d + \log\log T} + 1)^{1/2}}{\sqrt{\beta}d(t+1)} \right) \tag{244}$$

With a direct derivation from (244), one can show that

$$\overline{\Delta}_t = \prod_{s=t_0}^{t} \frac{s - \frac{2}{d-1}}{s+1}\Delta_{t_0} + \sum_{s=t_0}^{t}\left(\prod_{v=s}^{t} \frac{v - \frac{2}{d-1}}{v+1} \right) \cdot O\left(\frac{1}{s+1} \cdot \frac{(\sigma\sqrt{d + \log\log T} + 1)^{1/2}}{\sqrt{\beta}d} \right). \tag{245}$$

To upper bound $|\overline{\Delta}_t|$ in (245), we claim that for any $s < t$, it holds that

$$\prod_{v=s}^{t} \frac{v - \frac{2}{d-1}}{v+1} \lesssim \left(\frac{t}{s} \right)^{\frac{d+1}{d-1}}. \tag{246}$$

The proof of the claim is deferred to the last part of this section. With this result, one can upper bound $|\Delta_t|$ as

$$|\overline{\Delta}_t| \leq \left(\frac{t_4}{t} \right)^{\frac{d+1}{d-1}}|\overline{\Delta}_{t_0}| + \sum_{s=t_0}^{t}\left(\frac{s}{t} \right)^{\frac{d+1}{d-1}} \cdot O\left(\frac{1}{s} \cdot \frac{(\sigma\sqrt{d + \log\log T} + 1)^{1/2}}{\sqrt{\beta}d} \right)$$

$$\lesssim \left(\frac{\beta^8}{t\sigma^6} \right)^{\frac{d+1}{d-1}} + \frac{(\sigma\sqrt{d + \log\log T} + 1)^{1/2}}{\sqrt{\beta}d}, \tag{247}$$

where the second inequality holds as

$$\sum_{s=t_0}^{t} \frac{1}{s}\left(\frac{s}{t} \right)^{\frac{d+1}{d-1}} = \frac{1}{t^{\frac{d+1}{d-1}}}\sum_{s=t_o}^{t} s^{\frac{2}{d-1}} \lesssim 1.$$

As we note that $|b_t - b_\star| \leq \overline{\Delta}_t$, which leads to our conclusion that

$$\overline{\lambda}_t^2 = \beta^2 t \left(\frac{2}{d+1} + \overline{\Delta}_t \right), \tag{248}$$

which allows us to characterize $\overline{\lambda}_t$ as

$$\overline{\lambda}_t = \beta\sqrt{t} \cdot \sqrt{\frac{2}{d+1} + \overline{\Delta}_t} = \sqrt{\frac{2\beta^2 t}{d+1}} \cdot \sqrt{1 + \frac{d+1}{2}\overline{\Delta}_t}$$

$$= \sqrt{\frac{2\beta^2 t}{d+1}} \cdot (1 + \Delta_t) \tag{249}$$

where $\Delta_t \lesssim d\overline{\Delta}_t$, which implies that $|\Delta_t|$ is upper bounded as

$$|\Delta_t| \lesssim d \left(\frac{\beta^8}{t\sigma^6} \right)^{\frac{d+1}{d-1}} + \frac{(\sigma\sqrt{d + \log\log T} + 1)^{1/2}}{\sqrt{\beta}}, \tag{250}$$

Set $\widetilde{\lambda}_t = \sqrt{\frac{2\beta^2 t}{d+1}}$. Combining (249) and Proposition 4, one can show that

$$\frac{\lambda_{T,i}}{\widetilde{\lambda}_T} = \frac{\lambda_{T,i}}{\overline{\lambda}_T} \cdot \frac{\overline{\lambda}_T}{\widetilde{\lambda}_T}$$

$$= \left(1 + O\left(\frac{d(\sigma\sqrt{d + \log\log T} + 1)}{\beta} \right) \right) \cdot (1 + \Delta_t)$$

$$= 1 + O\left(d \left(\frac{\beta^8}{t\sigma^6} \right)^{\frac{d+1}{d-1}} + \frac{(\sigma\sqrt{d + \log\log T} + 1)^{1/2}}{\sqrt{\beta}} \right), \tag{251}$$

where the last equality holds as $\beta \gtrsim d^2(\sigma\sqrt{d + \log\log T} + 1)$. Hence, we establish the desired result.

Proof of Claim (246). Taking the logarithm on left side, one can see that

$$\log \left(\prod_{t=t_0}^{t_1} \frac{t - \frac{2}{d-1}}{t+1} \right) = \sum_{t=t_0}^{t_1} \log \left(1 - \frac{(d+1)/(d-1)}{t+1} \right)$$

$$\leq -\sum_{t=t_0}^{t_1} \frac{(d+1)/(d-1)}{t+1}$$

$$\leq -\int_{t=t_0}^{t_1} \frac{(d+1)/(d-1)}{t} dt + C$$

$$\leq -\frac{d+1}{d-1} \log \left(\frac{t_1}{t_0} \right) + C. \tag{252}$$

Therefore, one can show the desired result by taking exponential on both sides.

E   Proof of auxiliary lemmas

E.1   Proof of Lemma 1

From the definition of UCB score $\mathrm{UCB}_t(\boldsymbol{a})$, we note that,

$$\mathrm{UCB}_t(\boldsymbol{a}) = \langle \boldsymbol{a}, \widehat{\boldsymbol{\theta}}_{t-1} \rangle + \beta \cdot \sqrt{\boldsymbol{a}^\top \boldsymbol{\Lambda}_{t-1}^{-1} \boldsymbol{a}}$$

$$= \boldsymbol{a}^\top \widehat{\boldsymbol{\theta}}_t + \beta \cdot \|\boldsymbol{\Lambda}_{t-1}^{-1/2} \boldsymbol{a}\|_2. \tag{253}$$

Therefore, the action at time t can be equivalently characterized as

$$
\begin{aligned}
\boldsymbol{a}_t &= \arg \max_{\|\boldsymbol{a}\|_2=1} \boldsymbol{a}^\top \widehat{\boldsymbol{\theta}}_t + \beta \cdot \|\boldsymbol{\Lambda}_{t-1}^{-1/2} \boldsymbol{a}\|_2 \\
&= \arg \max_{\|\boldsymbol{a}\|_2=1} \boldsymbol{a}^\top \widehat{\boldsymbol{\theta}}_t + \beta \cdot \max_{\|\boldsymbol{w}\|_2=1} \boldsymbol{a}^T \boldsymbol{\Lambda}_{-1}^{-1/2} \boldsymbol{w} \\
&= \arg \max_{\|\boldsymbol{a}\|_2=1} \max_{\|\boldsymbol{w}\|_2=1} \boldsymbol{a}^\top (\widehat{\boldsymbol{\theta}}_t + \beta \cdot \boldsymbol{\Lambda}_{t-1}^{-1/2} \boldsymbol{w}).
\end{aligned} \tag{254}
$$

We may further note that

$$
\max_{\|\boldsymbol{a}\|_2=1} \max_{\|\boldsymbol{w}\|_2=1} \boldsymbol{a}^\top \left(\widehat{\boldsymbol{\theta}}_t + \beta \cdot \boldsymbol{\Lambda}_{t-1}^{-1/2} \boldsymbol{w}\right) = \max_{\|\boldsymbol{w}\|_2=1} \|\widehat{\boldsymbol{\theta}}_t + \beta \cdot \boldsymbol{\Lambda}_{t-1}^{-1/2} \boldsymbol{w}\|_2,
$$

with the equality holds if and only if $\boldsymbol{a} = \mathcal{P}(\widehat{\boldsymbol{\theta}}_t + \beta \cdot \boldsymbol{\Lambda}_{t-1}^{-1/2} \boldsymbol{w})$. Therefore, we note that

$$
(\boldsymbol{a}_t, \boldsymbol{w}_t) = \arg \max_{\|\boldsymbol{a}\|_2=1} \arg \max_{\|\boldsymbol{w}\|_2=1} \boldsymbol{a}^T \left(\widehat{\boldsymbol{\theta}}_t + \beta \cdot \boldsymbol{\Lambda}_{t-1}^{-1/2} \boldsymbol{w}\right), \tag{255}
$$

which is equivalent to the formula expressed in the lemma.

E.2   Proof of Lemma 2

As (37) holds, $\widehat{\boldsymbol{\theta}}_t$ is concentrated close to the top eigenspace of $\boldsymbol{\Lambda}_t$, then one may control $\nu_{t,i}$ for $i \geq 2$ as follows,

$$
\nu_{t,i} = \langle \boldsymbol{v}_{t,i}, \widehat{\boldsymbol{\theta}}_t \rangle = \langle \boldsymbol{v}_{t,i}, \widehat{\boldsymbol{\theta}}_t - \boldsymbol{v}_{t,1} \rangle \leq h_t, \quad \forall\, i \geq 2, \tag{256}
$$

Furthermore, we note that as $\{\boldsymbol{v}_{t,1}, \ldots, \boldsymbol{v}_{t,d}\}$ forms a orthogonal basis, the following will hold for $\nu_{t,1}$,

$$
\sum_{i=2}^{d} \nu_{t,i}^2 = \sum_{i=2}^{d} \langle \boldsymbol{v}_{t,i}, \widehat{\boldsymbol{\theta}}_t - \boldsymbol{v}_{t,1} \rangle^2 \leq \left\|\widehat{\boldsymbol{\theta}}_t - \boldsymbol{v}_{t,1}\right\|_2^2 \leq h_t^2, \tag{257}
$$

which directly yields that

$$
\nu_{t,1} \geq \sqrt{1 - \sum_{i=2}^{d} \nu_{t,i}^2} \geq 1 - h_t^2. \tag{258}
$$

implying that $\widehat{\boldsymbol{\theta}}_t$ indeed concentrates around the top eigenspace. For the next step, we recall our previous decomposition of the actions taken by LinUCB:

$$
\begin{aligned}
\boldsymbol{w}_t &= \arg \max_{\|\boldsymbol{w}\|_2=1} \left\|\widehat{\boldsymbol{\theta}}_t + \beta \cdot \boldsymbol{\Lambda}_t^{-1/2} \boldsymbol{w}\right\|_2, \\
\boldsymbol{a}_t &= \mathcal{P}\left(\widehat{\boldsymbol{\theta}}_t + \beta \boldsymbol{\Lambda}_t^{-1/2} \boldsymbol{w}_t\right),
\end{aligned}
$$

Therefore, one may conclude that

$$
\|\boldsymbol{a}_t - \widehat{\boldsymbol{\theta}}_t\|_2 = \left\|\mathcal{P}(\widehat{\boldsymbol{\theta}}_t + \beta \boldsymbol{\Lambda}_t^{-1/2} \boldsymbol{w}_t) - \mathcal{P}(\widehat{\boldsymbol{\theta}}_t)\right\|_2 \leq \left\|\beta \boldsymbol{\Lambda}_t^{-1/2} \boldsymbol{w}_t\right\|_2 \leq \frac{\beta}{\sqrt{\lambda_{t,d}}}. \tag{259}
$$

Therefore, one can conclude that

$$
\begin{aligned}
\|\boldsymbol{\xi}_t\|_2 &\leq \|\boldsymbol{a}_t - \widehat{\boldsymbol{\theta}}_t\|_2 = O\left(\frac{\beta}{\sqrt{\lambda_{t,d}}}\right), \\
\alpha_t &= \sqrt{1 - \|\boldsymbol{\xi}_t\|_2^2} = 1 - O\left(\frac{\beta^2}{\lambda_{t,d}}\right). \tag{260}
\end{aligned}
$$

Since we have

$$\kappa_{t,i} = \langle \boldsymbol{v}_{t,i}, \boldsymbol{a}_t \rangle = \alpha_t \langle \boldsymbol{v}_{t,i}, \widehat{\boldsymbol{\theta}}_t \rangle + \langle \boldsymbol{v}_{t,i}, \boldsymbol{\xi}_t \rangle = \alpha_t \nu_{t,i} + \langle \boldsymbol{v}_{t,i}, \boldsymbol{\xi}_t \rangle, \tag{261}$$

we can calculate $\kappa_{t,1}$ as follows:

$$\begin{aligned}
\kappa_{t,1} &= \alpha_t \nu_{t,1} + \langle \boldsymbol{v}_{t,1}, \boldsymbol{\xi}_t \rangle \\
&= \left(1 - O\left(\frac{\beta^2}{\lambda_{t,d}} \right) \right) \cdot \left(1 - O(h_t^2) \right) + O\left(\frac{\beta}{\sqrt{\lambda_{t,d}}} \right) \cdot O(h_t) \\
&= 1 - O\left(\frac{\beta^2}{\lambda_{t,d}} \right) - O(h_t^2).
\end{aligned} \tag{262}$$

We can also characterize $\kappa_{t,i}$ for $i \geq 2$ as

$$\begin{aligned}
\kappa_{t,i} = \alpha_t \nu_{t,i} + \langle \boldsymbol{v}_{t,i}, \boldsymbol{\xi}_t \rangle &= \left(1 - O\left(\frac{\beta^2}{\lambda_{t,d}} \right) \right) \cdot O(h_t) + O\left(\frac{\beta}{\sqrt{\lambda_{t,d}}} \right) \\
&= O(h_t) + O\left(\frac{\beta}{\sqrt{\lambda_{t,d}}} \right),
\end{aligned} \tag{263}$$

which concludes the desired result.

E.3 Proof of Lemma 3

From (35), one can note that

$$\begin{aligned}
\sum_{i=1}^d \left(\nu_{t,i} + \frac{\beta w_{t,i}}{\sqrt{\lambda_{t,i}}} \right)^2 &\geq \sum_{i=1}^d \nu_{t,i}^2 + \left(\nu_{t,d} + \frac{\beta \cdot \mathrm{sign}(\nu_{t,d})}{\sqrt{\lambda_{t,d}}} \right)^2 \\
&\geq \sum_{i=1}^d \nu_{t,i}^2 + \frac{\beta^2}{\lambda_{t,d}}.
\end{aligned} \tag{264}$$

Define "small eigenvalues" to be the set of eigenvalues smaller than $C_1 \lambda_{t,d}$ for some constant C_1, and define the threshold to be $k_t = \max\{k : \lambda_{t,k} > C_1 \lambda_{t,d}\}$, then one can show that

$$\begin{aligned}
\sum_{i=1}^d \left(\nu_{t,i} + \frac{\beta w_{t,i}}{\sqrt{\lambda_{t,i}}} \right)^2 &\leq \left(\sqrt{\sum_{i=1}^d \nu_{t,i}^2} + \sqrt{\sum_{i=1}^d \frac{\beta^2 w_{t,i}^2}{\lambda_{t,i}}} \right)^2 \\
&\leq \left(\sqrt{\sum_{i=1}^d \nu_{t,i}^2} + \beta \sqrt{\sum_{i=1}^{k_t} \frac{w_{t,i}^2}{\lambda_{t,i}} + \sum_{i=k_t+1}^d \frac{w_{t,i}^2}{\lambda_{t,i}}} \right)^2 \\
&\leq \left(\sqrt{\sum_{i=1}^d \nu_{t,i}^2} + \beta \sqrt{\sum_{i=1}^{k_t} \frac{w_{t,i}^2}{C_1 \lambda_{t,d}} + \sum_{i=k_t+1}^d \frac{w_{t,i}^2}{\lambda_{t,d}}} \right)^2 \\
&\leq \left(\sqrt{\sum_{i=1}^d \nu_{t,i}^2} + \beta \sqrt{\frac{1}{C_1 \lambda_{t,d}} \left(1 - \sum_{i=j_t+1}^d w_{t,i}^2 \right) + \frac{1}{\lambda_{t,d}} \cdot \sum_{i=k_t+1}^d w_{t,i}^2} \right)^2 \\
&= \left(\sqrt{\sum_{i=1}^d \nu_{t,i}^2} + \frac{\beta}{\sqrt{\lambda_{t,d}}} \cdot \sqrt{\frac{1}{C_1} \left(1 + (C_1 - 1) \sum_{i=k_t+1}^d w_{t,i}^2 \right)} \right)^2.
\end{aligned} \tag{265}$$

Combining (264) and (265), we note that

$$\frac{\beta}{\sqrt{\lambda_{t,d}}} \leq \frac{2\sqrt{\sum_{i=1}^{d} \nu_{t,i}^2} \cdot \sqrt{\frac{1}{C_1}\left(1 + (C_1 - 1)\sum_{i=k_t+1}^{d} w_{t,i}^2\right)}}{1 - \frac{1}{C_1}\left(1 + (C_1 - 1)\sum_{i=k_t+1}^{d} w_{t,i}^2\right)}. \tag{266}$$

Based on Assumption 3, we note that

$$\frac{\beta/\sqrt{\lambda_{t,d}}}{\sqrt{\sum_{i=1}^{d} \nu_{t,i}^2}} \geq c/c_0.$$

Therefore, one can show that

$$\frac{2\sqrt{\frac{1}{C_1}\left(1 + (C_1 - 1)\sum_{i=k_t+1}^{d} w_{t,i}^2\right)}}{1 - \frac{1}{C_1}\left(1 + (C_1 - 1)\sum_{i=k_t+1}^{d} w_{t,i}^2\right)} \geq c/c_0, \tag{267}$$

which further implies that

$$\frac{1}{C_1}\left(1 + (C_1 - 1)\sum_{i=k_t+1}^{d} w_{t,i}^2\right) \geq \frac{c^2}{8c_0^2}. \tag{268}$$

As we set $C_1 = \frac{16c_0^2}{c^2} - 1$, we note that

$$\sum_{i=k_t+1}^{d} w_{t,i}^2 \geq \frac{C_1 c^2/8 - 1}{C_1 - 1} = \frac{1 - \frac{c^2}{8}}{\frac{16}{c^2} - 2} = \frac{c^2 c_0^2}{16}. \tag{269}$$

We then consider the projection of action on the set of "small eigenvalues". Recall that $\boldsymbol{a}_t = \sum_{i=1}^{d} \kappa_{t,i} \boldsymbol{v}_{t,i}$, where $\kappa_{t,i}$ can be expressed as

$$\kappa_{t,i} = \frac{\nu_{t,i} + \frac{\beta w_{t,i}}{\sqrt{\lambda_{t,i}}}}{\sqrt{\sum_{j=1}^{d}\left(\nu_{t,j} + \frac{\beta w_{t,j}}{\sqrt{\lambda_{t,j}}}\right)^2}},$$

which implies that the square summation of $\kappa_{t,i}$ can be upper bounded as

$$\begin{aligned}
\sum_{i=k_t+1}^{d} \kappa_{t,i}^2 &= \frac{\sum_{i=k_t+1}^{d}\left(\nu_{t,i} + \frac{\beta w_{t,i}}{\sqrt{\lambda_{t,i}}}\right)^2}{\sum_{j=1}^{d}\left(\nu_{t,j} + \frac{\beta w_{t,j}}{\sqrt{\lambda_{t,j}}}\right)^2} \\
&\geq \frac{\frac{\beta^2}{\lambda_{t,d}} \cdot \sum_{i=j_t+1}^{d} w_{t,i}^2}{\left(\sqrt{\sum_{j=1}^{d} \nu_{t,j}^2} + \frac{\beta}{\sqrt{\lambda_{t,d}}} \cdot \sqrt{\frac{1}{C_1}\left(1 + (C_1 - 1)\sum_{i=j_t+1}^{d} w_{t,i}^2\right)}\right)^2} \\
&\geq \frac{\frac{\beta^2}{\lambda_{t,d}} \cdot \frac{c^2 c_0^2}{16}}{\left(\sqrt{\sum_{j=1}^{d} \nu_{t,j}^2} + \frac{\beta}{\sqrt{\lambda_{t,d}}} \cdot \frac{c}{4}\right)^2} \\
&\geq \frac{c^4 c_0^2}{25}, \tag{270}
\end{aligned}$$

when c is set small enough, which concludes our claim when we set $C_1 = 16c_0^2/c^2 - 1$ and $C_2 = c^4/25$.

E.4 Proof of Lemma 4

We rewrite the problems as follows. Denote

$$\alpha_i = \frac{\beta}{\sqrt{\lambda_{t,i}}}, \quad b_i^{(1)} = \alpha_i \nu_{t,i}, \quad b_i^{(2)} = \alpha_1 \mathbf{1}(i = 1).$$

Then the $\boldsymbol{w}_t^\star$ is the solution of

$$f_1(\boldsymbol{w}) = \sum_{i=1}^d 2\alpha_i b_i^{(1)} w_i + \alpha_i^2 w_i^2,$$

while $\widetilde{\boldsymbol{w}}_t^\star$ is the solution of

$$f_2(\boldsymbol{w}) = \sum_{i=1}^d 2\alpha_i b_i^{(2)} w_i + \alpha_i^2 w_i^2.$$

The KKT condition gives the solution of both optimization problems to be

$$w_i^\star = \frac{b_i^{(1)}}{\mu_1 - \alpha_i^2}, \quad \widetilde{w}_i^\star = \frac{b_i^{(2)}}{\mu_2 - \alpha_i^2}, \tag{271}$$

where μ_1, μ_2 satisfies

$$\sum_{i=1}^d \left(\frac{b_i^{(1)}}{\mu_1 - \alpha_i^2} \right)^2 = 1, \quad \sum_{i=1}^d \left(\frac{b_i^{(2)}}{\mu_2 - \alpha_i^2} \right)^2 = 1.$$

A direct calculation yields that $\mu_1 > \alpha_d^2$ but $\mu_2 = \alpha_d^2$. We also note that $|b_i^{(1)}| \leq |b_i^{(2)}|$. Therefore, one obtain the following inequality

$$|w_{t,1}^\star| = \frac{|b_i^{(1)}|}{\mu_1 - \alpha_1^2} \leq \frac{|b_i^{(2)}|}{\mu_2 - \alpha_1^2} = |\widetilde{w}_{t,1}^\star|, \tag{272}$$

establishing the desired result.

E.5 Proof of Lemma 6

As we note that $\lambda_{t+1,1}$ is the largest root of the following equation

$$f(\lambda) = 1 + \sum_{i=1}^d \frac{\kappa_{t,i}^2}{\lambda_{t,i} - \lambda}. \tag{273}$$

Set $\widetilde{\lambda} = \lambda_{t,1} + \kappa_{t,1}^2$, then as $\lambda_{t,1} - \lambda_{t,i} \geq t/2$, for all $i \geq 2$, we have

$$f(\widetilde{\lambda}) = \sum_{i=2}^d \frac{\kappa_{t,i}^2}{\lambda_{t,i} - \lambda} = O(t^{-1}). \tag{274}$$

Furthermore, we also note that as $\kappa_{t,1}^2/2 \leq |\lambda - \lambda_{t,1}| \leq 3\kappa_{t,1}^2/2$, it holds that

$$f'(\lambda) = -\sum_{i=1}^d \frac{\kappa_{t,i}^2}{(\lambda_{t,i} - \lambda)^2} \leq -\frac{4}{9\kappa_{t,1}^2} + O(t^{-2}) \leq -\frac{4}{9} + O(t^{-2}). \tag{275}$$

Therefore, we conclude that

$$|\lambda_{t,1} - \widetilde{\lambda}| = O(t^{-1}), \tag{276}$$

implying that $\lambda_{t+1,1} = \lambda_{t,1} + \kappa_{t,1}^2 + O(t^{-1})$. We will then further characterize the growth rate by plugging in $\kappa_{t,1}$. From Lemma 2, one can show that

$$\kappa_{t,1} = \left(1 - O\left(\frac{\beta^2}{\lambda_{t,d}}\right)\right)\left(1 - O\left(\frac{\beta^2}{\lambda_{t,d}}\right)\right) + O\left(\frac{\beta}{\sqrt{\lambda_{t,d}}} \cdot \frac{\beta}{\sqrt{\lambda_{t,d}}}\right)$$

$$= 1 - O\left(\frac{\beta^2}{\lambda_{t,d}}\right), \tag{277}$$

Therefore, one can show that

$$\lambda_{t+1,1} = \lambda_{t,1} + 1 - O\left(\frac{\beta^2}{\lambda_{t,d}}\right) + O(t^{-1}) = \lambda_{t,1} + 1 - O\left(\frac{\beta^2}{\lambda_{t,d}}\right),$$

where the last equality holds as $\lambda_{t,d} \asymp \beta\sqrt{t}$.

E.6 Proof of Lemma 8

We begin with the following transformation. Let $\tilde{\boldsymbol{\lambda}} = \boldsymbol{\Lambda}^{1/2}\boldsymbol{\lambda}$ and $\tilde{\boldsymbol{\eta}} = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{\eta}$. Then $f(\boldsymbol{\lambda})$ can be rewritten as

$$f(\boldsymbol{\lambda}) = \tilde{\boldsymbol{\lambda}}^T\tilde{\boldsymbol{\eta}} - \frac{\sigma^2}{2}\tilde{\boldsymbol{\lambda}}^T\tilde{\boldsymbol{\lambda}} = -\frac{\sigma^2}{2}\left\|\tilde{\boldsymbol{\lambda}} - \frac{1}{\sigma^2}\tilde{\boldsymbol{\eta}}\right\|_2^2 + \frac{1}{2\sigma^2}\|\tilde{\boldsymbol{\eta}}\|_2^2,$$

with $\max_{\boldsymbol{\lambda}\in\mathbb{R}^d} f(\boldsymbol{\lambda}) = (2\sigma^2)^{-1}\|\tilde{\boldsymbol{\eta}}\|_2^2$. Then for any $\boldsymbol{\lambda}\in\mathcal{C}$, it holds that

$$-\frac{\sigma^2}{2}\left\|\tilde{\boldsymbol{\lambda}} - \frac{1}{\sigma^2}\tilde{\boldsymbol{\eta}}\right\|_2^2 + \frac{1}{2\sigma^2}\|\tilde{\boldsymbol{\eta}}\|_2^2 \geq \frac{1-\kappa^2}{2\sigma^2}\|\tilde{\boldsymbol{\eta}}\|_2^2,$$

which immediately implies

$$\left\|\tilde{\boldsymbol{\lambda}} - \frac{1}{\sigma^2}\tilde{\boldsymbol{\eta}}\right\|_2 \leq \frac{\kappa}{\sigma^2}\|\tilde{\boldsymbol{\eta}}\|_2. \tag{278}$$

Substituting $\boldsymbol{\lambda} = \boldsymbol{\Lambda}^{-1/2}\tilde{\boldsymbol{\lambda}}$ and $\boldsymbol{\eta} = \boldsymbol{\Lambda}^{1/2}\tilde{\boldsymbol{\eta}}$ into (278) yields the desired result.

E.7 Proof of Lemma 10

We set $a_n = A_n n^{-1/4}$, then it holds that

$$A_{n+1}^2(n+1)^{-1/2} \leq \left(\left(1 - \frac{1}{n}\right)A_n n^{-1/4} + \frac{B}{n^{5/4}}\right)^2 + \frac{C}{n^2}$$

$$= \left[\left(\left(1 - \frac{1}{n}\right)A_n + \frac{B}{n}\right)^2 + \frac{C}{n^{3/2}}\right] \cdot n^{-1/2}. \tag{279}$$

From the basic inequality that $(1 + 1/n)^{1/2} \leq 1 + 1/(2n)$, one can show that

$$A_{n+1}^2 \leq \left(1 + \frac{1}{2n}\right) \cdot \left[\left(\left(1 - \frac{1}{n}\right)A_n + \frac{B}{n^{5/4}}\right)^2 + \frac{C}{n^2}\right]$$

$$= \left(1 + \frac{1}{2n}\right) \cdot \left[\left(\left(1 - \frac{1}{n}\right) + \frac{B}{n^{5/4}A_n}\right)^2 + \frac{C}{n^2 A_n^2}\right] \cdot A_n^2$$

$$\leq \left(1 + \frac{1}{2n}\right) \cdot \left(1 - \frac{2}{n} + \frac{1}{n^2} + \frac{2B}{n^{5/4}A_n} + \frac{B^2}{n^{5/2}A_n^2} + \frac{C}{n^{3/2}A_n^2}\right) \cdot A_n^2. \tag{280}$$

Note that when
$$n \geq \max\left(12, \frac{4096B^4}{A_n^4}, \frac{12^{2/3}B^{4/3}}{A_n^{4/3}}, \frac{144C^2}{A_n^4}\right),$$

we have
$$\frac{1}{n^2} \leq \frac{1}{12n}, \quad \frac{2B}{n^{5/4}A_n} \leq \frac{1}{4n}, \quad \frac{B^2}{n^{5/2}A_n^2} \leq \frac{1}{12n}, \quad \frac{C}{n^{3/2}A_n^2} \leq \frac{1}{12n}.$$

then it holds that
$$
\begin{aligned}
A_{n+1}^2 &\leq \left(1 - \frac{1}{2n}\right) \cdot \left(1 - \frac{2}{n} + \frac{1}{12n} + \frac{1}{4n} + \frac{1}{12n} + \frac{1}{4n}\right) A_n^2 \\
&\leq \left(1 - \frac{1}{n}\right) A_n^2.
\end{aligned}
\tag{281}
$$

Now that we constrain on n such that $A_n \geq 4B$, then (281) holds as long as
$$n \geq \max\left(16, \frac{9C^2}{16B^4}\right),$$

Therefore, let $n_1 = \min\{n : A_{n+1} \leq 4B\}$, then for all $n_0 < n \leq n_1$, it holds that
$$A_n^2 \leq \prod_{i=n_0}^{n} \left(1 - \frac{1}{i}\right) \cdot A_{n_0}^2 \leq \frac{n_0}{n} \cdot A_{n_0}^2, \tag{282}$$

as $A_{n_1} \geq 4B$, (282) implies
$$\frac{n_0}{n_1} \cdot A_{n_0}^2 \geq (4B)^2,$$

which allows us to upper bound n_1 as
$$n_1 \leq \left(\frac{A_{n_0}}{4B}\right)^2 n_0 = \frac{a_{n_0}^2 n_0^{3/2}}{16B^2}. \tag{283}$$

We also note that the right hand side of (280) is monotone with respect to A_n, and when $A_n = 4B$, it holds that
$$A_{n+1}^2 \leq \left(1 - \frac{1}{n}\right) \cdot (4B)^2 = (4B)^2,$$

which implies that for any n such that $A_n \leq 4B$, we have
$$A_{n+1} \leq 4B. \tag{284}$$

Therefore, for all $n \geq n_1$, it holds that $A_n \leq 4B$. Combining with (283), we arrive at our final result.

References

Abbasi-yadkori, Y., Pál, D., and Szepesvári, C. (2011). Improved algorithms for linear stochastic bandits. In Shawe-Taylor, J., Zemel, R., Bartlett, P., Pereira, F., and Weinberger, K., editors, *Advances in Neural Information Processing Systems*, volume 24. Curran Associates, Inc.

Banerjee, D., Ghosh, A., Chowdhury, S. R., and Gopalan, A. (2023). Exploration in linear bandits with rich action sets and its implications for inference. In *International Conference on Artificial Intelligence and Statistics*, pages 8233–8262. PMLR.

Billingsley, P. (2013). *Convergence of probability measures*. John Wiley & Sons.

Bubeck, S., Cesa-Bianchi, N., et al. (2012). Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends® in Machine Learning*, 5(1):1–122.

Chen, X., Lee, J. D., Tong, X. T., and Zhang, Y. (2020). Statistical inference for model parameters in stochastic gradient descent.

Deshpande, Y., Javanmard, A., and Mehrabi, M. (2023). Online debiasing for adaptively collected high-dimensional data with applications to time series analysis. *Journal of the American Statistical Association*, 118(542):1126–1139.

Deshpande, Y., Mackey, L., Syrgkanis, V., and Taddy, M. (2018). Accurate inference for adaptive linear models. In *International Conference on Machine Learning*, pages 1194–1203. PMLR.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366a):427–431.

Dimakopoulou, M., Ren, Z., and Zhou, Z. (2021). Online multi-armed bandits with adaptive inference. *Advances in Neural Information Processing Systems*, 34:1939–1951.

Dwork, C., Feldman, V., Hardt, M., Pitassi, T., Reingold, O., and Roth, A. (2015). The reusable holdout: Preserving validity in adaptive data analysis. *Science*, 349(6248):636–638.

Fan, Y., Han, Y., Lv, J., Xu, X., and Zhou, Z. (2024). Precise asymptotics and refined regret of variance-aware ucb. *arXiv preprint arXiv:2412.08843*.

Golub, G. H. and Van Loan, C. F. (2013). *Matrix computations*. JHU press.

Guo, Y. and Xu, Z. (2025). Statistical inference for misspecified contextual bandits. *arXiv preprint arXiv:2509.06287*.

Hadad, V., Hirshberg, D. A., Zhan, R., Wager, S., and Athey, S. (2021). Confidence intervals for policy evaluation in adaptive experiments. *Proceedings of the national academy of sciences*, 118(15):e2014602118.

Halder, B., Pan, S., and Khamaru, K. (2025). Stable thompson sampling: Valid inference via variance inflation. *arXiv preprint arXiv:2505.23260*.

Han, Q., Khamaru, K., and Zhang, C.-H. (2024). Ucb algorithms for multi-armed bandits: Precise regret and adaptive inference. *arXiv preprint arXiv:2412.06126*.

He, J., Zhao, H., Zhou, D., and Gu, Q. (2023). Nearly minimax optimal reinforcement learning for linear markov decision processes.

Kalvit, A. and Zeevi, A. (2021). A closer look at the worst-case behavior of multi-armed bandit algorithms. In *NeurIPS 2021 (Spotlight)*.

Kausik, C., Tan, K., and Tewari, A. (2024). Leveraging offline data in linear latent bandits. *arXiv preprint arXiv:2405.17324*.

Khamaru, K. and Zhang, C.-H. (2024). Inference with the upper confidence bound algorithm. *arXiv preprint arXiv:2408.04595*.

Lai, T. L. (1987). Adaptive treatment allocation and the multi-armed bandit problem. *The annals of statistics*, pages 1091–1114.

Lai, T. L. and Robbins, H. (1985). Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics*, 6(1):4–22.

Lai, T. L. and Wei, C. Z. (1982). Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *The Annals of Statistics*, pages 154–166.

Lattimore, T. and Szepesvári, C. (2020). *Bandit algorithms*. Cambridge University Press.

Li, L., Chu, W., Langford, J., and Schapire, R. E. (2010). A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*, WWW '10, page 661–670. ACM.

Lin, L., Ying, M., Ghosh, S., Khamaru, K., and Zhang, C.-H. (2023). Statistical limits of adaptive linear models: low-dimensional estimation and inference. *Advances in Neural Information Processing Systems*, 36:15965–15990.

Nair, Y. and Janson, L. (2023). Randomization tests for adaptively collected data. *arXiv preprint arXiv:2301.05365*.

Niu, Z. and Ren, Z. (2025). Assumption-lean weak limits and tests for two-stage adaptive experiments. *arXiv preprint arXiv:2505.10747*.

Polyak, B. T. and Juditsky, A. B. (1992). Acceleration of stochastic approximation by averaging. *SIAM journal on control and optimization*, 30(4):838–855.

Robbins, H. (1952). Some aspects of the sequential design of experiments.

Sengupta, I. and Khamaru, K. (2024). Stable batched bandit: Optimal regret with free inference. OpenReview. ICLR 2025 submission.

Sherman, U., Cohen, A., Koren, T., and Mansour, Y. (2024). Rate-optimal policy optimization for linear markov decision processes.

Shi, L., Wang, J., and Wu, T. (2023). Statistical inference on multi-armed bandits with delayed feedback. In *International Conference on Machine Learning*, pages 31328–31352. PMLR.

Su, W. J. and Zhu, Y. (2023). Higrad: Uncertainty quantification for online learning and stochastic approximation. *Journal of Machine Learning Research*, 24(124):1–53.

Tan, K., Fan, W., and Wei, Y. (2025). Actor-critics can achieve optimal sample efficiency. *arXiv preprint arXiv:2505.03710*.

Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294.

van der Vaart, A. (2000). Asymptotic statistics. *Cambridge Books*.

Vershynin, R. (2018). *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press.

Waudby-Smith, I., Wu, L., Ramdas, A., Karampatziakis, N., and Mineiro, P. (2024). Anytime-valid off-policy inference for contextual bandits. *ACM/IMS Journal of Data Science*, 1(3):1–42.

Wu, W., Li, G., Wei, Y., and Rinaldo, A. (2024). Statistical inference for temporal difference learning with linear function approximation. *arXiv preprint arXiv:2410.16106*.

Wu, W., Wei, Y., and Rinaldo, A. (2025). Uncertainty quantification for markov chains with application to temporal difference learning. *arXiv preprint arXiv:2502.13822*.

Zhang, K. W., Janson, L., and Murphy, S. A. (2020). Inference for batched bandits: Highlighting non-normality of bandit estimates. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pages 9818–9829.

Zhang, K. W., Janson, L., and Murphy, S. A. (2021). Statistical inference with m-estimators on adaptively collected data. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34.